I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13593

Correcting for Misclassified Binary Regressors Using Instrumental Variables

Steven J. Haider
Melvin Stephens Jr.

AUGUST 2020

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13593

Correcting for Misclassified Binary Regressors Using Instrumental Variables

Steven J. Haider
Michigan State University and IZA

Melvin Stephens Jr.
University of Michigan and NBER

AUGUST 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

IZA DP No. 13593 AUGUST 2020

ABSTRACT

Correcting for Misclassified Binary Regressors Using Instrumental Variables*

Estimators that exploit an instrumental variable to correct for misclassification in a binary regressor typically assume that the misclassification rates are invariant across all values of the instrument. We show that this assumption is invalid in routine empirical settings. We derive a new estimator that is consistent when misclassification rates vary across values of the instrumental variable. In cases where identification is weak, our moments can be combined with bounds to provide a confidence set for the parameter of interest.

JEL Classification: C18, C26

Keywords: misclassification, measurement error, instrumental variables

Corresponding author:
Steven J. Haider
Department of Economics
Michigan State University
101 Marshall Hall
East Lansing, MI 48824
USA
E-mail: haider@msu.edu

* We would like to thank Dan Black, Andreas Hagemann, Gary Solon, Jeff Wooldridge, and seminar participants at the University of Chicago Harris School, the University of Michigan Labor Lunch, Michigan State University, the 2019 Society of Labor Economists meetings, and the Michigan-Michigan State-Western University Labo(u)r Day Conference for helpful comments and suggestions.

1 Introduction

It has long been recognized that measurement error is pervasive in applied economic research (e.g., Bound, Brown, and Mathiowetz 2001). When a continuous regressor is mis-measured, it is well-known that standard instrumental variables (IV) estimation will yield a consistent estimate if the measurement error is "classical," i.e., uncorrelated with the true regressor. However, the same approach does not work with a binary regressor because the measurement error associated with a misclassified binary variable will be negatively correlated with the true regressor. To address this issue, a number of empirical strategies have been developed to consistently estimate the impact of a misclassified binary regressor, many of which make use of an instrumental variable (e.g., Card 1996; Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000; Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007; Hu 2008; Chen, Hu, and Lewbel 2008a,b; Battistin, De Nadai, and Sianesi 2014; DiTraglia and Garcia-Jimeno 2019; Calvi, Lewbel, and Tomassi 2017; Yanagi 2019).

A key identifying assumption used in the prior literature is that the conditional probabilities of misclassifying the binary regressor, or the "misclassification rates," do not depend upon the value of the instrumental variable.[1] However, this seemingly innocuous assumption is, in fact, rather strong. For example, in the literature that examines the impact of Medicaid (e.g., Currie and Gruber 1996; Gross and Notowidigdo 2011), the observed binary indicator for Medicaid eligibility is constructed (primarily) by determining whether measured household income is below a given threshold (e.g., a state-specific threshold that depends on family size) and changes in the threshold across states and over time are used to construct an instrumental variable. Intuitively, misclassification of Medicaid eligibility will be driven by those with true income that is relatively close to the eligibility threshold and, thus, are more likely to have their measured income fall on the wrong side of the threshold. Since the share of households close to the threshold, and therefore most at risk of being misclassified, can vary as the threshold moves through the income distribution, the misclassification rates of Medicaid eligibility can vary across values of the instrumental variable. Consistent with this intuition, we present evidence from multiple empirical settings illustrating that misclassification rates vary in this way.

[1] Specifically, the misclassification rates are the probability of misclassifying the binary regressor conditional on the true value of the indicator.

We refer to misclassification rates that vary with the instrumental variable as "varying misclassification" in contrast to the standard "fixed misclassification" assumption used in the prior literature.[2] Prior instrumental variables-related solutions that assume fixed misclassification are inconsistent when binary regressors suffer from varying misclassification. In addition, prior research has shown that standard IV estimation will overestimate the parameter of interest under the fixed misclassification assumption and have suggested using the IV estimator as an upper bound (Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000).[3] In contrast to this earlier analysis, standard IV estimation can produce either overestimates or underestimates with varying misclassification. Because prior estimators are inconsistent and IV estimation does not yield an upper bound with varying misclassification, it is important to develop new estimation methods that account for this type of misclassification.

In this paper, we present an estimator that consistently estimates the impact of a misclassified, binary regressor, even when the misclassification rate varies across instrument values. Identification requires a discrete instrumental variable that takes on three or more values. Our key identifying assumption, which we discuss in detail below, places a restriction on the form of the varying misclassification that is suggested by our empirical examples. It is straightforward to include covariates in our framework, and we can allow the misclassification rates to vary with the covariates. We also provide a new lower bound for the impact of the misclassified binary regressor when instruments are available.

The estimating equations that identify our model can be implemented with Generalized Method of Moments (GMM). However, the GMM estimator can fail to converge when the instrument is weak and/or the true impact of the binary indicator is small (i.e., close to zero).[4] To address this concern, we add parameter bounds via inequality constraints to our set of moment conditions.

[2] Although earlier papers allow misclassification rates to vary with other observable characteristics, they do not allow the misclassification rates to vary with the instrumental variable. Recent exceptions are Ura (2018) and Yanagi (2019). Below we show how to extend our results to allow for misclassification that varies both with other observable characteristics *and* the instrumental variable.

[3] Ura (2018) develops an alternative upper bound based on differences in the joint distribution of the outcome and the misclassified regressor between groups with different values of the instrument. Nguimkeu, Denteh, and Tchernis (2019) show that the IV estimator will not be an upper bound in the case of endogenous misreporting. The OLS regression of an outcome on a misclassified binary regressor yields a corresponding lower bound (Aigner 1973; Bollinger 1996).

[4] DiTraglia and Garcia-Jimeno (2019) highlight this issue in context of misclassified binary regressors.

While the methods found in Andrews and Soares (2010) can be applied to construct confidence sets for the model parameters after including these inequality moments, we apply the method of Bugni, Canay, and Shi (2017) to construct the marginal confidence set for the parameter on the binary indicator, which is typically the parameter of interest.

A variety of methodological approaches have been developed to estimate the impact of a mismeasured binary regressor. Our paper is most closely related to those which point identify the parameter of interest using an instrumental variable (Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007; Hu 2008; DiTraglia and Garcia-Jimeno 2019); however, these papers require fixed misclassification. Another set of papers relies on the availability of multiple measures of the misclassified treatment (Card 1996; Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000; Battistin, De Nadai, and Sianesi 2014), but still require assumptions analogous to fixed misclassification. Recent papers require both an instrumental variable and at least one additional variable which provides information on the measurement error in order to achieve identification (Calvi, Lewbel, and Tomassi 2017; Yanagi 2019), whereas our approach only requires a single instrumental variable. Several other papers have utilized restrictions on higher order moments to achieve identification (e.g., Chen, Hu, and Lewbel 2008a,b; DiTraglia and Garcia-Jimeno 2019), but these approaches introduce arguably strong assumptions on the distribution of the errors which we are able to circumvent.[5] Finally, we contribute a new lower bound on the parameter of interest, following a long tradition in this literature (Aigner 1973; Bollinger 1996; Black, Berger, and Scott 2000; Kreider et al 2012; Ura 2018).

Two recent papers allow for varying misclassification. Ura (2018) bounds the impact of a mismeasured treatment allowing for general forms of misclassification. While his main result assumes that misclassification is independent of the instrument, he also provides an additional result which allows misclassification to be correlated with the instrument. Yanagi (2019) point identifies the impact of a mismeasured binary regressor, but requires both an instrumental variable and an additional covariate that provides identifying information on the measurement error process. In this paper we are able to point identify the impact of a mismeasured binary regressor in the presence of varying misclassification using only a single discrete valued instrument.

[5] DiTraglia and Garcia-Jimeno (2019) use both an instrumental variable and and higher order moment restrictions.

The paper is set out as follows. In the next section, we provide motivating empirical examples to illustrate the importance of allowing for misclassification rates to vary with the instrumental variable. We then develop our main theoretical result. After discussing model estimation, we provide a Monte Carlo simulation study to illustrate our main findings. The final section concludes.

2 Examples of Varying Misclassification

To illustrate the potential importance of varying misclassification, we present examples using multiple data sources. We first use data from the March 1973 Current Population Survey (CPS) that is linked both to tax returns filed with the Internal Revenue Service (IRS) (Social Security Administration 2005).[6] The earnings data are for the prior calendar year, 1972. A drawback to the IRS earnings data is that these are taken directly from tax filings, which only contains a single, combined earnings measure for all household members. As such, we restrict our sample to men and women whose marital filing status on their tax return is listed as single and for whom the match between the CPS and the administrative sources is designated as "good." We further restrict the analysis to men and women who report in the CPS that they are privately employed, have positive IRS and CPS earnings, and have non-imputed CPS earnings. We use the sample weights, which account for both selection into the CPS as well as the match between the CPS and the administrative records.[7]

The top panel of Figure 1 shows the probability density function of both (log) administrative and self-reported earnings in 1972 for our sample of 8,031 single men and women. We treat the IRS earnings data as the individual's "true" earnings E^*, and the CPS earnings report as the mis-measured earnings E that researchers typically have available for their analysis. The two distributions appear quite similar throughout, with the most noticeable differences occurring in the middle of the distributions.

For this example, we construct a hypothetical program eligibility measure T^*, which equals one for all individuals with true earnings E^* at or below a threshold c, and equals zero otherwise. At

[6] The data also contain Social Security (SS) Earnings records. We use the IRS earnings data because the available SS earnings records are top-coded at the annual ceiling for earnings subject to the SS tax, which censors earnings for nearly half of privately employed men. The IRS earnings data as well as self-reported earnings in the CPS are top-coded at $50,000, which affects less than one-half of one percent of privately employed men in the data.

[7] More details on the CPS linked data and sample selection criteria are provided in Appendix Section A.1.

the threshold c_1, represented by the leftmost vertical line in the top panel of Figure 1, 39.1 percent of individuals have earnings at or below threshold, and thus are coded as $T^* = 1$.[8] Of course, the researcher will typically only have access to the analogous measure T, which is determined by whether E is at or below c. With the threshold set a c_1, $T = 1$ for 40.4 percent of individuals.

As we will discuss in detail below, the misclassification rates are key inputs determining the extent to which standard OLS and IV estimators are biased due to misclassification. These conditional probabilities are defined as $\alpha_0 = P[T = 1|T^* = 0]$ and $\alpha_1 = P[T = 0|T^* = 1]$. When the threshold is at c_1, we calculate that $\alpha_0 = 0.068$ and $\alpha_1 = 0.072$. At c_2, the second threshold shown in the top panel of Figure 1, $T^* = 1$ for 87.2 percent of sample members while $T = 1$ for 87.7 percent. However, the misclassification rates found at c_2, $\alpha_0 = 0.149$ and $\alpha_1 = 0.016$, are markedly different from those found at c_1.

The bottom panel of Figure 1 plots the misclassification rates for numerous values of c, placing $P[T^* = 1]$ on the horizontal axis. Notice that α_1 is decreasing in $P[T^* = 1]$. Intuitively, when a larger fraction of the population is eligible for the program ($T^* = 1$), the share of eligible individuals who are close to the eligibility threshold is smaller. For analogous reasons, α_0 increases with $P[T^* = 1]$.

In a typical program eligibility setting, such as the Medicaid example mentioned in the Introduction, changes in the eligibility threshold are the basis for an instrumental variable. For example, Currie and Gruber (1996) treat the change in a state's eligibility threshold as an exogenous source of variation for the fraction of individuals in the state who have access to program benefits. However, as with the example shown in Figure 1, the misclassification rates for the observed, binary regressor T will systematically vary as these thresholds change. Such changes in the misclassification rates run counter to the fixed misclassification assumption needed for consistency of the estimation methods found in the prior literature.

In addition, the distinction between varying and fixed misclassification has important implications for interpreting the results from standard IV estimation. The standard IV estimator for the impact of T^* on Y when using Z as an instrument is equivalent to $COV(Y, Z)/COV(T^*, Z)$. If we only

[8] This threshold, $log(c_1) = 7.65$, is approximately \$2,100 which was the poverty line for a single, unrelated individual in 1972.

have access to the mis-measured regressor, T, the standard IV estimator is $COV(Y,Z)/COV(T,Z)$. Because the numerators of these two quantities are identical, any inconsistency that arises by applying IV estimation to T instead of T^* is due to differences in the denominators of these expressions.

To gain insight into the differences between the denominators of these IV estimators, consider the case where Z is a binary instrument. In this case, $COV(T^*,Z) = P[T^* = 1|Z = 1] - P[T^* = 1|Z = 0]$ and $COV(T,Z) = P[T = 1|Z = 1] - P[T = 1|Z = 0]$. Thus, the relationship between the IV estimator with T versus T^* is determined by how $P[T = 1]$ changes relative to $P[T^* = 1]$ between the two instrument values.

Figure 2 plots the relationship between $P[T = 1]$ and $P[T^* = 1]$ under the assumption of fixed misclassification (the line labeled "Hypothetical Fixed Misclassification" in the top panel) and for what is observed in the actual data (the line labeled "Actual $P[T = 1]$" in the bottom panel).[9] For the fixed misclassification case shown in the top panel, the slope of the line is constant and always less than one (i.e., flatter than the 45 degree line). As such, a given change in a threshold c will always result in a smaller change in $P[T = 1]$ than the corresponding change in $P[T^* = 1]$.[10] Thus, when using T instead of T^*, the standard IV estimator with fixed misclassification will have a smaller denominator ($COV(T,Z) < COV(T^*,Z)$) and therefore will overestimate the impact of T^* on Y.[11]

However, as we see in the bottom panel of Figure 2, the actual relationship between $P[T = 1]$ and $P[T^* = 1]$ is non-linear. Around $P[T^* = 1] = 0.5$, the slope of this curve is less than one, similar to what occurs in the fixed misclassification case. On the other hand, at both small and large values of $P[T^* = 1]$, the change in $P[T = 1]$ is *larger* than the corresponding change in $P[T^* = 1]$. In these ranges, applying the standard IV estimator to T will underestimate the impact of T^* on Y. Thus, contrary to the received wisdom from the prior literature, IV estimation may either overestimate or underestimate the true parameter when misclassification varies with the instrument.

We find comparable evidence of varying misclassification if we construct similar hypothetical

[9] For the hypothetical fixed misclassification case, we use $\alpha_0 = 0.068$ and $\alpha_1 = 0.072$, which are the values corresponding the c_1 threshold shown in Figure 1.

[10] Since $P[T = 1] = \alpha_0 + (1 - \alpha_0 - \alpha_1)P[T^* = 1]$, the slope of this line will be less than one unless T is perfectly measured. In addition, under the standard assumption that $\alpha_0 + \alpha_1 < 1$, which is routinely made in this literature, T and T^* will be positively correlated and the slope of this line will exceed zero.

[11] See, e.g., Kane, Rouse, and Staiger 1999; Black, Berger, and Scott 2000; Ura 2018

program eligibility measures using wage data from matched employer-employer reports in the January 1977 CPS.[12] These data have been used previously to examine misclassification in union status (Freeman 1984; Card 1996) and measurement error in wages (Mellow and Sider 1983; Angrist and Krueger 1999). The top panel of Figure 3 demonstrates that the misclassification rates are varying as the threshold moves through the wage distribution. Interestingly, the bottom panel of Figure 3 shows that changes in $P[T = 1]$ are nearly identical to changes in $P[T^* = 1]$, suggesting that IV using T rather than T^* will yield roughly consistent estimates throughout much of the wage distribution.

While our motivating examples thus far have focused on earnings, similar patterns can emerge in starkly different settings. For example, consider a binary variable based on height and weight data from the 1999-2016 waves of the Continuous National Health and Nutrition Examination Survey (Continuous NHANES). Survey participants first undergo a detailed in-home interview during which self-reported height (in inches) and weight (in pounds) are recorded. Participants then undergo a detailed medical exam about two weeks later at a mobile examination center. During this examination, each participant's height (in centimeters) and weight (in kilograms) are measured by survey staff. Analogous to our use of the earnings data above, we create a binary variable indicating whether an individual's height or weight is less than a particular threshold using both the self-reported data and the clinically-obtained data.[13]

Figure 4 shows height results separately for men and women.[14] The misclassification rates for men, shown in the top left panel, vary with $P[T^* = 1]$ in ways that closely mirror those found in the previous two examples. Moreover, the relative movements in $P[T = 1]$ and $P[T^* = 1]$ are comparable to the first example in that slope of the line is less than one near the middle of the distribution but exceeds one towards the ends of the distribution. Such patterns indicate that the IV estimate may either overestimate or underestimate the true impact of a change in the dichotomous variable on the outcome of interest, depending upon which points in the distribution

[12] We thank David Card, Hank Farber, and Alan Krueger for making this data available to us.

[13] We limit the sample to men and women between the ages of 25 and 54, inclusive. For each outcome, we restrict the sample to individuals who have both self-reported and measured data. Our height samples have 11,004 men and 12,100 women while our weight sample has 10,694 men and 11,756 women. We did not include weight observations for the small fraction of individuals who were flagged for being weighed while wearing their clothing.

[14] Appendix Figure A1 present an analogous figure for weight.

are being compared.

The height patterns for women are different. As shown in the top right panel of Figure 4, while α_0 (α_1) is increasing (decreasing) with $P[T^* = 1]$ as in the prior examples, α_0 exceeds α_1 much further to the left in the distribution. As shown in the bottom right panel, underreporting of height occurs, on average, throughout the distribution as we find that $P[T = 1]$ always exceeds $P[T^* = 1]$. We find that the relationship between $P[T = 1]$ and $P[T^* = 1]$ moves much further from the 45 degree line than in the other examples, which implies much larger deviations between the change in $P[T = 1]$ and the change in $P[T^* = 1]$. At various points in the distribution, IV will generate larger underestimates and overestimates as compared to the previous examples.[15]

Overall, these examples yield two main implications. First, the misclassification rates vary with the instrument (here, the threshold), a finding that is at odds with the predominate fixed misclassification assumption in the prior literature.[16] Second, changes in $P[T = 1]$ may be greater than, less than, or very close to changes in $P[T^* = 1]$ when moving between thresholds. As such, the IV estimator is no longer guaranteed to yield an upper bound for the true impact of T^* on the outcome of interest. These findings are in contrast to the standard, fixed misclassification case.

3 The Model and Identification

The equation relating the outcome Y to the true, binary regressor $T^* = \{0, 1\}$ is[17]

$$Y = \gamma + \beta T^* + \epsilon \tag{1}$$

However, instead of observing T^*, we observe T, which is subject to misclassification, i.e., T is also a binary variable in which either $T = T^*$ or $T = 1 - T^*$.

[15] The patterns found for weight, as shown in Appendix Figure A1 are broadly comparable to those for height with regards to the relationships between $P[T^* = 1]$ and the misclassifications and in that the relationship between $P[T = 1]$ and $P[T^* = 1]$ deviates further from the 45 degree line for women than for men.

[16] While our motivating examples focus on misclassification rates varying when a threshold varies, the results we present in the next section apply more generally to whenever misclassification rates vary with an instrument. See Bound, Brown, and Mathiowetz (2001) for an exhaustive review of validation studies for labor market outcomes, including for evidence of when misclassification rates vary.

[17] Our key identification result can be extended to a non-parametric regression framework such as that used in Mahajan (2006) and Lewbel (2007). We adopt a parametric regression framework for parsimony of notation and for ease in adding covariates.

Let Z be a discrete-valued instrumental variable with $J+1$ values, $j = 0, 1, \ldots, J$. To allow for varying misclassification, define the misclassification rates, α_{0j} and α_{1j}, $j = 0, 1, \ldots, J$, as

$$\begin{aligned} \alpha_{0j} &= P[T = 1|T^* = 0, Z = j], \\ \alpha_{1j} &= P[T = 0|T^* = 1, Z = j] \end{aligned}$$

Defining $p_j^* = P[T^* = 1|Z = j]$, it follows that $p_j = P[T = 1|Z = j] = (1 - \alpha_{1j})\, p_j^* + \alpha_{0j}\left(1 - p_j^*\right) = \alpha_{0j} + (1 - \alpha_{0j} - \alpha_{1j})\, p_j^*$.

We make use of the following two assumptions for the majority of our analysis.

Assumption 1 *The following conditions are assumed to hold*

i. *There is a discrete instrumental variable, Z, with (at least) three values, i.e., $J \geq 2$*

ii. $E\left[Y|Z = j\right] = \gamma + \beta p_j^*,\ \forall j$

iii. $0 < p_j^* < 1,\ \forall j$*;* $p_j^* \neq p_k^*,\ \forall\ j \neq k$

iv. $E\left[Y|T, T^*, Z\right] = E\left[Y|T^*, Z\right]$

v. $\alpha_{0j} + \alpha_{1j} < 1,\ \forall j$

Assumption 1.ii is the usual instrumental variable exclusion condition and Assumption 1.iii is the usual instrumental variable relevance condition. The last two parts of Assumption 1 are standard in the prior literature (e.g., Frazis and Loewenstein 2003; Mahajan 2006; Lewbel 2007; DiTraglia and Garcia-Jimeno 2019). Assumption 1.iv states that there is no additional independent information contained in the mis-measured regressor once the true value of the regressor is known, which is often referred to as "non-differential" measurement error. This assumption is analogous to the standard classical measurement error assumption that the measurement error is random.[18] Assumption 1.v requires that there is not "too much" misclassification. With this restriction, the OLS regression of Y on T using observations with the same value of Z will yield an estimated coefficient on the mis-measured binary regressor that is of the same sign as β.

[18] A recent exception is Nguimkeu, Denteh, and Tchernis (2019) which allows for differential measurement error.

Assumption 2 *For all j, $E\left[Y|T^{*}=t,Z=j\right]=E\left[Y|T^{*}=t\right]$*

Assumption 2 requires the expected value of the outcome Y for a given value of the true regressor T^{*} to be the same across all values taken on by the instrument Z. The same assumption is made by Mahajan (2006) when using a binary instrumental variable to identify the impact of a misclassified binary regressor when there is fixed misclassification. Similar assumptions have been made for other proposed estimators in this literature (e.g., Kane, Rouse, and Staiger 1999, Black, Berger, and Scott 2000; Frazis and Loewenstein 2003). More recently, this assumption has been proposed as a test of no selection bias in the literature on instrumental variables (Black et al. 2015; Brinch, Mogstad, and Wiswall 2017). While T^{*} is exogenous under Assumption 2, we relax this assumption below to admit some forms of endogeneity.

3.1 Identification

For a discrete instrumental variable Z that takes on exactly three discrete values, Assumption 1.ii yields three equations based on the exclusion restriction

$$\begin{aligned} E\left[Y|Z=0\right] &= \gamma+\beta p_0^{*} \\ E\left[Y|Z=1\right] &= \gamma+\beta p_1^{*} \\ E\left[Y|Z=2\right] &= \gamma+\beta p_2^{*} \end{aligned} \tag{2}$$

Assumption 2 yields two equations for any pair of instrument values, j and k

$$E\left[Y|T^{*}=1,Z=j\right]=E\left[Y|T^{*}=1,Z=k\right] \tag{3}$$

$$E\left[Y|T^{*}=0,Z=j\right]=E\left[Y|T^{*}=0,Z=k\right] \tag{4}$$

Equations (3) and (4) are based on unobserved quantities. For a given value of Z, the expected value of the outcome Y conditional on T is a weighted average of these unknown quantities, with the weights based on the misclassification rates. As shown in Appendix Section A.2, these relationships can be inverted to yield expressions for the unknown quantities in (3) and (4) in terms of observed quantities and the misclassification rates. The resulting expressions are

$$E\left[Y|T^{*}=1,Z=j\right]=\frac{COV\left(Y,T|Z=j\right)}{p_{j}-\alpha_{0j}}+E\left[Y|Z=j\right] \tag{5}$$

$$E\left[Y|T^{*}=0,Z=j\right]=E\left[Y|Z=j\right]-\frac{COV\left(Y,T|Z=j\right)}{1-p_{j}-\alpha_{1j}} \tag{6}$$

where $COV\left(Y,T|Z=j\right)$ is the covariance of Y and T among those observations where $Z=j$.

Subtracting equation (4) from equation (3) and then inserting equations (5) and (6) into the resulting expression yields (as shown in Appendix Section A.3)[19]

$$\frac{1}{\left(1-\alpha_{0j}-\alpha_{1j}\right)}\cdot\frac{COV\left(Y,T|Z=j\right)}{p_{j}^{*}\left(1-p_{j}^{*}\right)}=\frac{1}{\left(1-\alpha_{0k}-\alpha_{1k}\right)}\cdot\frac{COV\left(Y,T|Z=k\right)}{p_{k}^{*}\left(1-p_{k}^{*}\right)} \tag{7}$$

For a discrete instrument with three values, these substitutions provide five equations: three equations found in (2) and two additional equations by applying (7) to two pairs of instrument values. However, eleven unknown parameters appear in these five equations: β, γ, the three p_{j}^{*}, the three α_{0j}, and the three α_{1j}. Noting that misclassification rates do not appear in (2) and only appear in (7) as the sum $\alpha_{0j}+\alpha_{1j}$, we can instead view the problem as containing five equations with eight parameters: β, γ, the three p_{j}^{*}, and the three sums $\alpha_{0j}+\alpha_{1j}$.[20] Moving to an instrument with more discrete values will not solve this identification issue because each additional instrument value adds two equations but also two parameters: another p_{j}^{*} and another sum $\alpha_{0j}+\alpha_{1j}$.

Instead, we make an assumption regarding the misclassification rates. As we discussed above, the fixed misclassification assumption made in the prior literature is too restrictive because the underlying misclassification rates likely vary across instrument values. We instead propose an alternative assumption.

Assumption 3 *The sum of the misclassification rates is constant, i.e.,* $\alpha_{0j}+\alpha_{1j}=\overline{\alpha}$, $\forall j$.

Assumption 3 allows both α_{0j} and α_{1j} to vary across instrument values j. Importantly, this

[19] By first subtracting equation (4) from equation (3) yields $E\left[Y|T^{*}=1,Z=j\right]-E\left[Y|T^{*}=0,Z=j\right]=E\left[Y|T^{*}=1,Z=k\right]-E\left[Y|T^{*}=0,Z=k\right]$. An alternative approach, discussed in Section 3.2, is to replace Assumption 2 with this expression (as in Lewbel (2007)).

[20] Alternatively, identification cannot be achieved by using the definition of p_{j} to replace p_{j}^{*} since the α_{0j} and α_{1j} will no longer enter the moments as a sum in the absence of p_{j}^{*}.

assumption nests the typical fixed misclassification assumption, making our estimator applicable under the conditions assumed in previous papers. In addition, this assumption is consistent with the empirical settings of Section 2: the misclassification rates move in opposite directions as $P[T^* = 1]$ changes while the magnitude of their sum, $\alpha_{0j} + \alpha_{1j}$, is relatively constant for much of the range of $P[T^* = 1]$.

Figure 5 plots the sum of the misclassification rates, along with bootstrapped 95% confidence intervals, for the empirical examples presented in Section 2. For the 1973 CPS-IRS matched earnings data and the 1977 CPS employer-employee matched wage data, shown in the two leftmost panels of Figure 5, the sum $\alpha_{0j}+\alpha_{1j}$ is relatively constant for much of the interior of the range of $P[T^* = 1]$. The sums of the misclassification rates for both male and female heights from the Continuous NHANES data, shown in two center panels of the Figure, are roughly constant across the vast majority of values of $P[T^* = 1]$. The sum of the weight misclassification rates, shown in the two rightmost panels, is not as flat across the range as is the case for height; however, the sums for weight only exhibit a relatively small change for much of its range. Overall, we interpret the results in Figure 5 as indicating that Assumption 3 is supported by these empirical applications, especially when $P[T^* = 1]$ takes on values towards the middle of its range. Importantly, for our proposed estimator to be consistent, Assumption 3 only needs to hold at the values spanned by the p_j^* in a given application, as opposed to across the entire range of $P[T^* = 1]$.

Equation (7) is greatly simplified under Assumption 3 because replacing $\alpha_{0j} + \alpha_{1j}$ with $\overline{\alpha}$ leads the misclassification rates to drop out of the expression to yield

$$\frac{COV\left(Y, T|Z = j\right)}{p_j^*\left(1 - p_j^*\right)} = \frac{COV\left(Y, T|Z = k\right)}{p_k^*\left(1 - p_k^*\right)} \tag{8}$$

Thus, under Assumptions 1 through 3, equation (2) holds for all $J + 1$ instrument values while the model generates J independent equations based on equation (8) using pairs of instrument values. If only a two-valued instrument were available, we would have three equations but four unknown parameters (β, γ, p_0^*, and p_1^*). However, an instrument with three values will generate five unique equations, three due to equation (2) and two due to equation (8), which allows us to identify the five parameters β, γ, p_0^*, p_1^*, and p_2^*.

Based on these equations, we now state the main theorem of the paper.

Theorem 1 *The parameters β, γ, p_0^*, p_1^*, p_2^*, and $\overline{\alpha}$ are identified under Assumptions 1, 2, and 3.*

We provide the detailed proof in Appendix Section A.4. In brief, we first combine and simplify the five moments based on (2) and (8) to generate two non-linear equations in p_0^* and p_2^* (without loss of generality). The resulting system of equations has four solutions. However, two of the solutions for the (p_0^*, p_2^*) pair are $(0,0)$ and $(1,1)$, which are ruled out by Assumption 1.iii. For the remaining two solutions, we are able to take each (p_0^*, p_2^*) pair to solve for p_1^*, β, and γ.

The two solutions for β are of equal magnitude, but opposite in sign

$$\beta = \pm \frac{\sqrt{D_3}}{\Big((C_2 - C_1)\,\Delta_{10} + (C_0 - C_1)\,\Delta_{21} \Big)} \tag{9}$$

where $\Delta_{jk} = E\left[Y|Z=j\right] - E\left[Y|Z=k\right]$, $C_j = COV\left(Y,T|Z=j\right)$, $D_0 = C_0\Delta_{21}^2$, $D_1 = C_1\Delta_{20}^2$, $D_2 = C_2\Delta_{10}^2$, and $D_3 = D_0^2 + D_1^2 + D_2^2 - 2D_0D_1 - 2D_0D_2 - 2D_1D_2$.

To determine the sign of β, we show in the appendix that under Assumption 2

$$1 - \overline{\alpha} = \frac{COV\left(Y,T|Z=j\right)}{\beta p_j^* \left(1 - p_j^*\right)} \tag{10}$$

which follows from the OLS estimator of Y on T^* for a given value of the instrument j. Because β appears on the right-hand side of (10) and it only differs in sign across the two solutions per (9), one solution will have $\overline{\alpha} < 1$ and the other solution will have $\overline{\alpha} > 1$.[21] Since $\overline{\alpha}$ must be less than 1 by Assumption 1.v, we can then choose the appropriate solution.

3.2 Relaxing the Exogeneity Assumption

The estimator derived above can be applied to settings with a weaker exogeneity assumption.

Assumption 4 $E\left[Y|T^*=1,Z=j\right] - E\left[Y|T^*=0,Z=j\right] = E\left[Y|T^*=1,Z=k\right] - E\left[Y|T^*=0,Z=k\right]$ *for all j,k.*

[21] For the two solutions that we cannot immediately rule out by Assumption 1.iii, the estimators for a given p_j^* sum to one across the two solutions. In other words, the estimator for p_j^* for the first solution equals $1 - p_j^*$ for the second solution. Thus, the product $p_j^*\left(1 - p_j^*\right)$, which appears in the denominator of (10), is the same across both solutions.

In contrast to Assumption 2, which requires the level of the expected value of Y to be the same across all values of the instrument for a given value of T^*, Assumption 4 only requires the difference in the expected value of Y between the two values of T^* to be the same across all values of the instrument. A similar assumption is made by Lewbel (2007) in constructing an estimator for a misclassified binary regressor with fixed misclassification. Assumption 4 allows T^* to be endogenous under some (strong) conditions such as those found in Olsen's (1981) selection model (see Brinch, Mogstad, and Wiswall 2017).[22]

Corollary 1.1 *The magnitude of β, but not the sign, is identified under Assumptions 1i.-iv., 3, and 4.*

Inserting equations (5) and (6) into Assumption 4 and simplifying the resulting expression yields equation (7). Along with the remaining assumptions in Corollary 1.1, we can again generate equations (2) and (8) for two pairs of instrument values. Thus, following the proof of Theorem 1, we can recover β, γ, p_0^*, p_1^*, and p_2^* for the two sets of solutions. However, unlike in Theorem 1, we are unable to recover $\overline{\alpha}$ and therefore cannot sign β under the above assumptions.[23] We can still use this result, however, to test the null hypothesis of there being no impact of T^* on Y.

One possibility for identifying the sign of β is to impose a tighter restriction on the maximum amount of measurement error allowed. The standard restriction imposed when using the fixed misclassification assumption is that $\alpha_0 + \alpha_1 < 1$. Under this condition, the sign of the OLS estimator, when applied to observations with the same value of Z, does not change when using T in place of T^*. This condition also implies that the sign of the IV (Wald) estimator, when comparing across any pair of instrument values, is invariant to using T in place of T^*. However, the analogous assumption in the varying misclassification case, $\alpha_{0j} + \alpha_{1j} < 1$ (Assumption 1.v), only guarantees that the OLS estimator does not change sign, but does not guarantee the IV estimator will have the same sign when using either T or T^*.

Assuming that the sign of the IV estimator does not change when using T instead of T^* allows us to identify the sign of β.

[22] In addition to Assumption 4, Lewbel requires T^* to be (conditionally) exogenous.

[23] Assumption 3 allows us to eliminate the misclassification rates from (8) and recover β, γ, p_0^*, p_1^*, and p_2^*. However, without an additional assumption, such as the stronger restrictions imposed by Assumption 2, we cannot identify $\overline{\alpha}$ with just the system of equations generated by equations (2) and (8).

Assumption 5 *$sgn\left(p_j - p_k\right) = sgn\left(p_j^* - p_k^*\right)$ $\forall\ j,k$*

This assumption is consistent with Figures 2b and 3b because $P[T=1]$ in an increasing function of $P[T^*=1]$.

Corollary 1.2 *β, γ, p_0^*, p_1^*, and p_2^* are identified under Assumptions 1, 3, 4, and 5.*

As we show in Appendix Section A.4, we can solve for β, γ, p_0^*, p_1^*, and p_2^* for each of the two solutions by Corollary 1.1. For each candidate p_j^* solution, the competing solution is $1-p_j^*$ as these two solutions sum to one. In particular, the difference between any p_j^* and p_k^* is equal and opposite in sign across the two solutions. Thus, the Wald estimators implied by equation (2) can be written as

$$\beta = \frac{\Delta_{jk}}{p_j^* - p_k^*} \tag{11}$$

where $\Delta_{jk} = E\left[Y|Z=j\right] - E\left[Y|Z=k\right]$. Since Δ_{jk} is a known quantity, the sign of β depends on the sign of $p_j^* - p_k^*$. Thus, by invoking Assumption 5, we can immediately sign β. We are unable, however, to identify $\overline{\alpha}$ without making further assumptions.[24]

4 Estimation and Inference

We begin by discussing estimation of our model using Generalized Methods of Moments (GMM). We also discuss the inclusion of additional covariates in the model and in the estimation procedure. However, as we will make clear below, the GMM estimator will not always converge when β is small and/or the instruments are weak. As such, we discuss how to estimate our model by incorporating bounds on β via inequality moments and implementing the methods of Andrews and Soares (2010) and Bugni, Canay, and Shi (2017) to generate confidence sets that are robust to the issues raised below.

4.1 GMM Estimation

The estimating equations in (2) and (8) that we use in our identification proof are conditional on instrument values. We obtain unconditional moment conditions for GMM estimation by simply

[24] As shown in Appendix Section A.6, we are able to identify $\overline{\alpha}$ if, in addition to assumptions 1, 3, and 4 we also assume that the error term in equation 1 is homoskedastic.

rescaling these equations. The unconditional moment analogous to equation (2) is

$$(Z_j/\pi_j) \cdot \left(Y - \gamma - \beta p_j^*\right) = 0 \tag{12}$$

where $Z_j = 1(Z = j)$ with $1(\cdot)$ being the indicator function, $\pi_j = P[Z = j]$, and we suppress the i subscripts for the outcome Y and the indicator Z_j.

The unconditional moment analogous to equation (8) is

$$(Z_j/\pi_j) \cdot \left[\frac{YT - Y \cdot p_j}{p_j^*\left(1 - p_j^*\right)}\right] - (Z_k/\pi_k) \cdot \left[\frac{YT - Y \cdot p_k}{p_k^*\left(1 - p_k^*\right)}\right] = 0 \tag{13}$$

where we make use of the fact that $COV\left(Y, T|Z = j\right) = E\left[YT|Z = j\right] - E\left[Y|Z = j\right] \cdot E\left[T|Z = j\right]$.

These unconditional moments introduce additional parameters to be estimated. For each instrument value, we must estimate $p_j = P[T = 1|Z = j]$ and $\pi_j = P[Z = j]$. However, because $\sum_{j=0}^{K} \pi_j = 1$, we only need to estimate J of these $J + 1$ probabilities.

Letting $W' = (Y, T, Z)$ (and continuing to suppress the i subscript), our unconditional moment conditions are thus given by $E[g(W, \boldsymbol{\delta})] = 0$, where

$$g(W, \boldsymbol{\delta}) = \begin{pmatrix} (Z_j/\pi_j) \cdot \left(Y - \gamma - \beta p_j^*\right) \\ (Z_j/\pi_j) \cdot \left[\frac{YT - Y \cdot p_j}{p_j^*\left(1 - p_j^*\right)}\right] - (Z_k/\pi_k) \cdot \left[\frac{YT - Y \cdot p_k}{p_k^*\left(1 - p_k^*\right)}\right] \\ (Z_j/\pi_j) \cdot (T - p_j) \\ Z_j - \pi_j \end{pmatrix} \tag{14}$$

where $\boldsymbol{\delta} = (\beta, \gamma, \pi_j^*, \pi_j, p_j)$ is the vector of unknown parameters. When the instrument takes on three discrete values, this vector contains ten moment conditions: three moments for the first row of (14), two moments for the second row, three moments for the third row, and two moments for the final row (recall we do not need to estimate one of the π_j's). As $\boldsymbol{\delta}$ contains ten parameters, the system is exactly identified. Instruments with more than three discrete values will yield an over-identified system. Given our assumptions and our identification results, standard GMM results can be applied, providing us with consistency and asymptotic normality of $\hat{\boldsymbol{\delta}}$ (e.g., see Wooldridge

2010).

The moment conditions in (14) are applicable under either Assumption 2 or Assumption 4, our two alternatives regarding the exogeneity of T^*. If we use Assumption 5 then we can identify the sign of β with this set of moments, regardless of which exogeneity assumption we use. However, under Assumption 2 we can also identify $\overline{\alpha}$ using equation (10). While this equation can be solved after estimating (14), we can also include this additional moment in the system of equations and simultaneously estimate $\overline{\alpha}$. The empirical moment that corresponds to (10) is

$$(Z_j/\pi_j) \cdot \left[(1-\overline{\alpha}) \cdot \left(\beta p_j^* \left(1-p_j^*\right)\right) - (YT - Y \cdot p_j)\right] = 0 \tag{15}$$

Since this moment yields an identical estimate for $\overline{\alpha}$ for each instrument value j, in estimation we only use one moment condition based on equation (15) for an arbitrary instrument value j.

4.2 Including Covariates

We present two approaches to including covariates. The first approach is a straightforward method to incorporate both continuous and discrete covariates. However, this approach requires that the misclassification rates do not vary with the additional covariates (although these rates can vary with the instrument). The second approach allows for the misclassification rates to vary with the covariates as well as the instrument, but requires discrete covariates.

To include the covariates, we specify the equation of interest to be

$$Y = \gamma + \beta T^* + \boldsymbol{\psi} \boldsymbol{X} + \epsilon \tag{16}$$

where $\boldsymbol{X}$ is a vector of exogenous covariates (i.e., ϵ is uncorrelated with each element of $\boldsymbol{X}$) and that the classification error in T^* does not vary with $\boldsymbol{X}$ (e.g., $P[T=1|T^*=0, Z=j, \boldsymbol{X}] = P[T=1|T^*=0, Z=j]$).

As shown in Appendix Section A.5, after updating our assumptions to condition on $\boldsymbol{X}$, we can derive moment conditions analogous to (2) and (8) that account for the additional covariates. The moment condition analogous to equation (2) is

$$E[Y|Z=j,\boldsymbol{X}]-\gamma-\beta p_j^*-E[\boldsymbol{\psi}\boldsymbol{X}|Z=j,\boldsymbol{X}]=0 \tag{17}$$

while the moment condition analogous to (8) is

$$\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}-\frac{COV\left(\boldsymbol{\psi}\boldsymbol{X},T|Z=j\right)}{p_j^*\left(1-p_j^*\right)}=\frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)}-\frac{COV\left(\boldsymbol{\psi}\boldsymbol{X},T|Z=k\right)}{p_k^*\left(1-p_k^*\right)} \tag{18}$$

For estimation, the empirical analogs of these moments can substituted for the first two lines in (14). The moments corresponding to p_j and π_j in (14) are unaffected by the inclusion of X and do not need to be updated to reflect the changes in the first two moment equations. However, the vector of coefficients $\boldsymbol{\psi}$ now enter the system of equations. Fortunately, we can pre-estimate $\boldsymbol{\psi}$ because regressing Y on T and $\boldsymbol{X}$, while using Z as an instrument for T, will yield a consistent estimate of $\boldsymbol{\psi}$.[25] After replacing $\boldsymbol{\psi}$ with its pre-estimated value in (17) and (18), we can apply GMM although we would need to adjust for using pre-estimated parameters (Newey 1984).

Our second approach allows for the misclassification rates and the other parameters to vary with discrete covariates by expanding the estimating equations (e.g., Mahajan 2006, Lewbel 2007). For example, suppose that there is a dichotomous exogenous covariate V. We can allow the moment conditions, (2) and (8), along with our equations for p_j and π_j (and (10) if included) to depend on V, thereby allowing all parameters to vary with V. This expanded set of moments also can be estimated using GMM. In addition, we can impose restrictions on these parameters, e.g., β does not vary with V, that can be readily tested.

4.3 Bounds on β

There are multiple methods available for bounding the impact of a misclassfied regressor. As shown in Aigner (1973), $|\hat{\beta}^{OLS}|<|\beta|$ when T^* is exogenous, where $\hat{\beta}^{OLS}$ is the coefficient on T in the OLS regression of Y on T using the full sample of observations. Bollinger (1996) shows that $\hat{\beta}^{OLS}$ is a lower bound when considering only information in the first and second moments of Y and

[25] We discuss the consistency $\boldsymbol{\psi}$ when estimated this way in Appendix Section A.5.

T.[26] When T^* is endogenous, the regression of Y on T, the reduced form equation for the 2SLS estimator, yields a lower bound (e.g., Ura 2018).

With an instrument, we are able to tighten the lower bound as compared to OLS, and we can do so under weaker assumptions than are required above for point identification. Specifically, we only need a discrete instrument that takes on at least two values, and we can drop Assumption 3.

Proposition 1 *For a discrete valued instrument Z that takes on at least two values, under Assumptions 1.ii-v. and 2*

i. if $\hat{\beta}^{OLS} > 0$, then $\hat{\beta}^{OLS} \leq \hat{\beta}^{LB,+} \leq \beta$

ii. if $\hat{\beta}^{OLS} < 0$, then $\hat{\beta}^{OLS} \geq \hat{\beta}^{LB,-} \geq \beta$

where

$$\begin{aligned}\hat{\beta}^{LB,+} &= \max_Z\{E\left[Y|T=1, Z=j\right]\} - \min_Z\{E\left[Y|T=0, Z=j\right]\} \\ \hat{\beta}^{LB,-} &= \min_Z\{E\left[Y|T=1, Z=j\right]\} - \max_Z\{E\left[Y|T=0, Z=j\right]\}\end{aligned}$$

The proof is provided in Appendix Section A.7. To briefly sketch the proof, without loss of generality, suppose $\beta > 0$ in which case $\hat{\beta}^{OLS} = E\left[Y|T=1\right] - E\left[Y|T=0\right] > 0$. Since $E\left[Y|T=1\right]$ is a weighted average of the $E\left[Y|T=1, z=j\right]$, for at least one instrument value, k, we will have $E\left[Y|T=1, z=k\right] \geq E\left[Y|T=1\right]$. If more than one instrument value meets this criteria, assign k to be the largest $E\left[Y|T=1, z=k\right]$. Analogously, let $E\left[Y|T=0, z=l\right]$ be the smallest value with $E\left[Y|T=0, z=l\right] \leq E\left[Y|T=0\right]$. The difference between these two quantities is at least as large as $\hat{\beta}^{OLS}$ but will not exceed β.

As noted earlier, the IV estimator no longer yields an upper bound with varying misclassification. Bollinger (1996) provides an upper bound when T^* is exogenous that does not require an instrumental variable. Ura (2018) derives an upper bound when T^* is endogenous under the assumption of fixed misclassification and also provides an extension which yields an upper bound when the misclassification rates vary with the instrument.

[26] In situations when a second misclassified variable, $\tilde{T}$ where either $\tilde{T} = T^*$ or $\tilde{T} = 1 - T^*$, is available, Black, Berger, and Scott (2000) use this information to construct a tighter lower bound than $\hat{\beta}^{OLS}$.

4.4 Inference with Weak Identification

There is an important caveat to implementing GMM. Inspection of the equation for β found in (9) shows that it yields a real solution as long as the term $D3$ is non-negative. As is further shown in the proof of Theorem 1 in Appendix Section A.4, $D3$ must be strictly positive as the solution to the system of equations for the p_j^* terms have $D3$ in the denominator. Although $D3$ is positive in the population, except when $\beta = 0$ (see Section A.4.1), it may be negative when replaced with its sample analog. Such situations are more likely to occur when at least one of the following holds: i) β is close to zero, ii) there are small differences between p_j^* values, and iii) the sample size is relatively small. Under these conditions, the GMM estimator is weakly identified and may not converge in finite samples.[27]

To overcome this issue, we combine our moment conditions with moment inequalities for β in order to construct confidence sets for the parameters of interest. In addition to our own lower bound discussed above, numerous choices for both the lower and upper bound on β are available in the literature (e.g., Aigner 1973; Bollinger 1996; Ura 2018). While inference can be performed using the moment inequality methods of Andrews and Soares (2010), their approach requires the joint confidence set for all parameters to be constructed. To focus on the marginal confidence set for β, the methods found in Bugni, Canay, and Shi (2017) can be applied as shown in our simulations below.

5 Monte Carlo Simulations

In this section we present Monte Carlo simulations to highlight multiple aspects of our estimator. For each simulation, we first randomly assign observations to one of three instrument values, $j = 0, 1, 2$, with assignment probabilities of 0.382, 0.236, and 0.382, respectively.[28] Anticipating multiple approaches we will employ below to construct the misclassified binary indicator T, we next generate a random variable $E^* \sim N(7.87, 1.19)$ to match the 1973 CPS/IRS administrative log earnings data

[27] DiTraglia and Garcia-Jimeno (2019) note a similar issue with their estimator.

[28] These were constructed from an underlying standard normal random variable where we assign $j = 0$ if the random variable is less than -0.3, $j = 2$ if the random variable is greater than or equal to 0.3, and $j = 1$ otherwise.

shown in Section 2.[29] We set the true, unobserved binary indicator T^* equal to one if E^* falls below an instrument value-specific cutoff where for our baseline simulations these cutoff values are chosen such that $p_0^* = 0.35$, $p_1^* = 0.50$, and $p_2^* = 0.65$. We then construct the outcome Y for each observation based on equation (1) by drawing the error term $\epsilon \sim N(0, 0.25)$ and setting the true value of γ equal to one.

Finally, we construct the observed but misclassified binary indicator T. For our baseline simulations, we select misclassification rates similar to those found in Figure 1b, but maintain Assumption 3 that $\alpha_{0j} + \alpha_{1j} = \overline{\alpha}$. We set $\overline{\alpha} = 0.130$ and allow α_{0j} (and thus α_{1j}) to vary with $Z = j$ where $\alpha_{00} = 0.055$, $\alpha_{01} = 0.070$, and $\alpha_{02} = 0.085$. We then draw the random variable $u \sim \text{uniform}[0, 1]$. For observations with $T^* = 0$ and instrument value j, we set $T = 1$ if $u < \alpha_{0j}$ and $T = 0$ otherwise. Similarly, for observations with $T^* = 1$ and instrument value j, we set $T = 0$ if $u < \alpha_{1j}$ and $T = 1$ otherwise.

Table 1 illustrates that the GMM estimator may be affected by weak identification when β is (relatively) close to zero and/or the sample size is small (see Section 4.4).[30] The table presents results from using GMM to estimate the model for multiple combinations of β and the sample size N, using 1,000 iterations for each combination. Panel A of Table 1 shows the fraction of iterations in which the GMM estimator successfully converges for each combination of β and N.[31] Consistent with the conditions for weak identification, we see that GMM estimator is less likely to converge as β approaches zero and/or as the sample size decreases. When $N = 1,000$, we find that GMM fails to converge for at least some fraction of iterations for every value of β. However, when $N = 100,000$ the GMM estimator always converges with the exception of when $\beta = 0$, the value of the parameter for which our estimator is not identified.

Even when GMM converges, weak identification yields imprecise estimates and raises concerns with inference as shown in the remaining Panels of Table 1. In Panel B we see that the average simulation point estimate when GMM converges is too large when $N = 1,000$. However, we find

[29] We specify E^* to have the same mean and variance as the log of 1973 administrative earnings.

[30] Another situation that can lead to weak identification is when there are small differences between the p_j^* values. We do not explore this aspect of weak identification in the simulations reported here.

[31] All GMM simulations are completed in MatLab. We obtain starting values using fminunc and its default settings. We obtain final parameter values using fminsearch, setting the search tolerance at 10^{-7} and the maximum iterations at 20,000.

that the estimates of β are well-centered when $N = 100,000$. Moreover, as shown in Panel C, GMM yields relatively larger standard errors when identification is weak, but as the sample size increases, the GMM estimates are quite precise. Finally, in Panel D, we see that the coverage rates for the 95% confidence intervals are as expected when $N = 100,000$, whereas coverage is worse when $N = 1,000$.[32]

In light of these findings, we present two sets of simulation results in the remainder of this section. In the first subsection, we present "large sample" results with $N = 100,000$ and $\beta = 1$ to demonstrate the performance of our estimator in the absence of the weak identification concerns. We focus on how our estimator and 2SLS perform under different assumptions regarding misclassification error. We also examine the robustness of our estimator when we relax Assumption 3 (i.e., that $\alpha_{0j} + \alpha_{1j}$ is constant) in a manner that mimics the examples from Section 2. In the second subsection, we present "small sample" results with $N = 1,000$ in which we examine inference when identification is weak. We use the methods of Bugni, Canay, and Shi (2017) to construct confidence sets for β, combining the moments from (14) with bounds on β.

5.1 "Large Sample" Simulations

Table 2 illustrates the performance of our estimator in large samples ($N = 100,000$). We maintain the same data generating process for the true parameters (Y, T^*, and Z) as in Table 1 and setting $\beta = 1$, but we vary the misclassification process (i.e., vary in how we construct T). In column (1) of Table 2, we present results when using "fixed misclassification" where we impose the baseline misclassification rates for the instrument value $j = 1$ ($\alpha_{01} = 0.070$ and $\alpha_{11} = 0.060$) across all instrument values. As noted above, the standard fixed misclassification assumption is a special case of our Assumption 3 and therefore our estimator is consistent in this setting. As expected, OLS is attenuated when using T as the regressor with a point estimate of 0.87 and a standard deviation (SD) across iterations of 0.002. The 2SLS estimator delivers 1.150 (SD=0.011), which is biased upwards by the amount that is given by the standard misclassification formula (i.e., $\beta/(1 - \alpha_0 - \alpha_1)$.[33] Our estimator is well-centered, delivering a mean estimate of 1.000 (SD=0.033).

[32] In simulations results shown in Appendix Tables A1-A3, we find that these patterns shown in Table 1 remain when we vary $\overline{\alpha}$.

[33] With $\beta = 1$, $\alpha_0 = 0.070$, and $\alpha_1 = 0.060$, the 2SLS estimator will converge to $1/(1 - 0.070 - 0.060) = 1.149$.

We next examine the performance of our estimator under varying misclassification. In column (2), we use the baseline varying misclassification rates from Table 1. For the 2SLS estimator, we obtain an average estimate of 1.032 (SD=0.009), which is noticeably smaller than what the 2SLS estimator would converge to under fixed misclassification (i.e., 1.149).[34] As before, our estimator is centered on the true value (1.003 with SD=0.033). In column (3), we increase the rate at which the misclassification rates vary with p^*, but continue to maintain $\overline{\alpha} = .13$. We now find that the average 2SLS estimate is *below* the true value of β (0.907 with SD=0.008). This result is consistent with our earlier analytical finding: under varying misclassification, the 2SLS estimator no longer provides an upper bound for β. Once again, our estimator provides is well-centered with a mean estimate of 1.00 (SD=.033).

In the last row of Table 2, we present the estimates of the lower bound we propose in Section 4.3 that is designed to improve upon OLS with T, which has been previously proposed as a lower bound in the literature (Aigner 1973; Bollinger 1996). Our lower bound is tighter than OLS regardless of whether the misclassification is fixed or varies. For example, in column 1, OLS with T delivers 0.870 (SD=0.002), whereas our lower bound is 0.928 (SD=0.003).

We next examine the robustness of our estimator to small deviations from Assumption 3 (i.e., that $\alpha_{0j} + \alpha_{1j}$ is constant). We do so in a situation where the true regressor T^* is generated from an underlying index, E^*, but we must construct the observed indicator T using the mis-measured index E. For our simulations, we construct E by adding a mean zero measurement error ν to the random variable E^*.[35] We then set the observed binary indicator T equal to one if E is less than the same instrument value-specific cutoff that we use to generate T^* from E^*. Appendix Figure A2 illustrates the misclassification rates resulting from using 500,000 simulated draws when constructing T in this way. The sum of the misclassification rates is U-shaped, reaches a minimum at $p^* = 0.5$, and increases most rapidly as p^* approaches both zero and one, very similar to what we observed in Figure 5 that was based on the actual data.

The last three columns of Table 2 present simulation results using this alternative approach for

[34] Given the distribution of observations across the three instrument values, the average misclassification rates for this column are the same as the fixed misclassification rates in column 1 (i.e., $\alpha_0 = 0.070$, and $\alpha_1 = 0.060$).

[35] Thus, E and E^* have the same mean. We select the variance of ν such that the variance of E equals 1.25, the same as the variance of the 1973 CPS/IRS self-reported log earnings data shown in Section 2.

constructing T. In column (4) we continue using the same values of p^* for each instrument value as in our baseline model. As shown in the table, the sum of the misclassification rates is almost identical across all three instrument values for these simulations and the average simulation estimate is very close to the true value (0.998 with SD=0.033). In column (5) we decrease the p^* values but maintain the same distance between instrument values in terms of probability ($p_0^* = 0.10$, $p_1^* = 0.25$, and $p_2^* = 0.4$). In doing so, the sum of the misclassification rates differs somewhat more than in the baseline case, while our estimator remains centered very close to the truth (1.006 with SD=0.046). Finally, in the last column of Table 2 we further shift down the p^* values ($p_0^* = 0.05$, $p_1^* = 0.20$, and $p_2^* = 0.35$). Although the maximum difference in the sum of the misclassification rates across instrument values is about 10%, our estimator continues to perform relatively well (1.012 with SD=0.052). Thus, for small deviations from Assumption 3, our estimator yields estimates of β that are very close to the true parameter value.

As we have discussed throughout, other papers have proposed methods to estimate the impact of a mismeasured binary regressor. In Appendix Table A4, we show three estimators that rely on fixed misclassification and a single instrumental variable (Frazis and Lowenstein 2003; Mahajan 2006; Lewbel 2007) are not well-centered in simulations which have varying misclassification. However, three estimators that instead rely on higher-order moments of the error term (Chen, Hu, and Lewbel 2008a, b; DiTraglia and Garcia-Jimeno 2019) are well-centered (see column (2) of Appendix Table A4) as the errors used in our simulations (which are normally distributed) satisfy the error term restrictions required by these estimators. However, as shown in the final three columns of Appendix Table A4 where we relax these higher moment error term assumptions, we find that these alternative methods yield estimates of β that systematically deviate from its true value while our estimator remains well-centered in all cases.

5.2 "Small Sample" Simulations

Much theoretical work in econometrics in recent years has been devoted to inference when parameters defined by a set of moment equalities and inequalities are weakly identified or even not point-identified at all (e.g., Chernozhukov, Hong and Tamer 2007; Romano and Shaikh 2008; Andrew and Soares 2010). These papers present techniques to construct *joint* confidence sets for the entire unknown parameter vector. As was our focus in the previous subsection, our interest is to draw inferences on our key parameter β. Thus, we make use of the related methods in Bugni, Canay, and Shi (2017; BCS hereafter) to construct a *marginal* confidence set for β alone. The basic idea of BCS is the same as papers such as Andrew and Soares (2010), except that BCS depart in their treatment of the non-focal parameters: after assuming a particular value for β, BCS propose to minimize the test statistic with respect to these other parameters and derives the distribution for this minimized test statistic. This method of inference is well-suited to our problem because it allows us to make use of the key moment conditions in (14) even when the parameters are weakly identified and to incorporate additional inequality moments to make use of bounds on β.

In Figure 6, we plot coverage curves for β (i.e., the plot of $1-\alpha$ for various nulls) using the BCS method for samples with $N = 1,000$ and for three different true values of β: Panel A for $\beta = 1$, Panel B for $\beta = 0.5$, and Panel $\beta = 0.25$. As noted above, in addition to the ten moment equalities in (14), our estimation procedure uses two moment inequalities to bound β: one using OLS as the lower bound and the other using the upper bound proposed in Bollinger (1996).[36] We also use two moment inequalities, based on equation (10), to bound $\overline{\alpha}$ between zero and one.[37] The coverage curves are derived by evaluating β at nulls at intervals of 0.025 (i.e., for Panel A, we compute $1-\alpha$ at 0.500, 0.525, 0.550, etc.). We follow the methods as laid out in BCS, with one modification: we pre-estimate the bounds, and then adjust the variance of the moments that use the pre-estimated bounds through a direct bootstrap procedure.[38]

Turning to the results, we see that this method yields coverage curves that include the true value and can reject the null $\beta = 0$ in each panel of Figure 6. Notice that each panel also includes two dotted, vertical lines which represent the (population) lower and upper bounds on β. As is clear, the coverage curves drop markedly just outside the bounds in each case, consistent with the

[36] In our simulations, we use OLS as the lower bound rather than the alternative lower bound we proposed above because OLS is very close to the true value, particularly for $\beta = 0.25$ in Panel C, and it has lower variance.

[37] The lower bound follows from $\overline{\alpha}$ being the sum of two probabilities and the upper bound follows from Assumption 1.v.

[38] Specifically, we compute a pre-estimated bound for each observation based on a bootstrap with replacement. Therefore, when we calculate the variance of the moments, the variability of the pre-estimated parameters is directly incorporated. This method of estimation has the benefits that it speeds processing time and that it allows us to make use of bounds that are not readily expressed as moments, as is the case with the Bollinger upper bound. We show in Appendix Figure A3 that this method to adjust for pre-estimation closely approximates the coverage curve when using OLS for the lower bound and the inverse of the reverse OLS regression for the upper bound, an upper bound that is analyzed in Klepper (1988) and readily expressed as a moment inequality.

inequality moments being the key moments to rejecting the null values. When $\beta = 1$ (Panel A), the implied 0.95 confidence set of $[0.88, 1.19]$ compares favorably to the implied 0.95 confidence set using GMM with $N = 100,000$ for the same data generating process ($[0.94, 1.07]$ based on column 2 of Table 2). The implied confidence sets in Panels B and C, for which identification is weaker because the true β is closer to 0, are somewhat wider.

6 Conclusion

In this paper, we extend the literature on estimation with a misclassified, binary regressor in an important dimension: we allow for the misclassification rates to vary with the instrumental variable. We first show that variability in the misclassification rates both arises in empirically relevant settings and overturns some of the key theoretical results on misclassification that are routinely relied upon. We then derive a new estimator that matches the conditions found in these empirical setting. We additionally extend our analytic results along several dimensions, including demonstrating how to include covariates, extending the main results to handle a specific form of endogeneity, and developing a new lower bound for β. Finally, we demonstrate with Monte Carlo evidence that our estimator can be applied in large and small sample settings and it continues to perform well when we allow the sum of the misclassification rates to vary in a manner that matches our motivating example.

One appealing feature of our key identifying assumption, $\alpha_{0j} + \alpha_{1j} = \overline{\alpha},\ \forall j$, is that it nests the standard fixed misclassification assumption. Alternative features of varying misclassification could be exploited in order to identify the model, such as parameterizing α_0 and/or α_1 as function(s) of p^*. Such an approach will almost certainly not nest the standard fixed misclassification model so researchers should be cautious when implementing alternatives along these lines. Future work might explore the fruitfulness of related modeling choices.

7 References

Aigner, Dennis J. (1973) “Regression with a binary independent variable subject to errors of observation,” *Journal of Econometrics*, 1:49-59.

Andrews, Donald W. K. and Gustavo Soares (2010) “Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection,” *Econometrica*, 78:119-157.

Angrist, Joshua D. and Alan B. Krueger (1999) “Empirical Strategies in Labor Economics,” in Orley Ashenfelter and David Card, eds., Handbook of Labor Economics, Vol 3A, (North-Holland, Amsterdam) 1277-1366.

Battistin, Erich, Michele De Nadai, and Barbara Sianesi (2014) “Misreported Schooling, Multiple Measures and Returns to Educational Qualifcations,” *Journal of Econometrics*, 181:136-150.

Black, Dan A., Mark C Berger, and Frank A Scott (2000) “Bounding Parameter Estimates with Nonclassical Measurement Error,” *Journal of the American Statistical Association*, 95(451):739-748.

Black, Dan A., Joonhwi Joo, Robert LaLonde, Jeffrey A. Smith, and Evan J. Taylor (2015) “Simple Tests for Selection Bias: Learning More from Instrumental Variables,” *IZA Discussion Paper No. 9346*.

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall (2017) “Beyond LATE with a Discrete Instrument,” *Journal of Political Economy*, 125(4): 985-1039.

Bollinger, Christopher R. (1996) “Bounding mean regressions when a binary regressor is mismeasured,” *Journal of Econometrics*, 73(2):387-399.

Bound, John, Charles C. Brown, and Nancy Mathiowetz (2001) “Measurement error in survey data,” in Handbook of Econometrics. Vol. 5, Elsevier, 3705-3843.

Bugni, Federico A., Ivan A. Canay, and Xiaoxia Shi (2017) “Inference for subvectors and other functions of partially identified parameters in moment inequality models,” *Quantitative Economics*, 8, 1-38.

Calvi, Rossella, Arthur Lewbel, and Denni Tommasi (2017) “LATE With Mismeasured or Misspecified Treatment: An Application To Women’s Empowerment in India”

Card, David (1996) “The Effect of Unions on the Structure of Wages: A Longitudinal Analysis,” *Econometrica*, 64(4):957-979.

Chen, Xiaohong, Yingyao Hu and Arthur Lewbel, (2008a) "A note on the closed-form identification of regression models with a mismeasured binary regressor," *Statistics & Probability Letters*, 78(12):1473-1479.

Chen, Xiaohong, Yingyao Hu and Arthur Lewbel, (2008b) "Nonparametric identification of regression models containing a misclassified dichotomous regressor without instruments," *Economics Letters*, 100(3):381-384.

Currie, Janet and Jonathan Gruber (1996) "Health Insurance Eligibility, Utilization of Medical Care, and Child Health," *The Quarterly Journal of Economics*, 111(2):431-466.

DiTraglia, Francis J. and Camilo Garcia-Jimeno (2019) "Identifying the Effect of a Mis-classified, Binary, Endogenous Regressor," *Journal of Econometrics*, 209:376-390.

Frazis, Harley and Mark A. Loewenstein (2003) "Estimating linear regressions with mismeasured, possibly endogenous, binary explanatory variables," *Journal of Econometrics*, 117(1):151-178.

Freeman, Richard B. (1984) "Longitudinal Analysis of the Effects of Trade Unions," *Journal of Labor Economics*, 2(1):1-26.

Gross, Tal and Matthew Notowidigdo (2011) "Health Insurance and the Consumer Bankruptcy Decision: Evidence from Medicaid Expansions," *Journal of Public Economics*, 95(7-8):767-778.

Hu, Yingyao (2008) "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144:27-61.

Kane, Thomas J., Cecilia Elena Rouse, and Douglas Staiger (1999) "Estimating Returns to Schooling When Schooling is Misreported," *National Bureau of Economic Research Working Paper No. 7235.*

Klepper, Steven (1988) "Bounding the Effects of Measurement Error In Regressions Involving Dichotomous Variables," *Journal of Econometrics*, 37:343-359.

Kreider, Brent, John V. Pepper, Craig Gundersen, and Dean Jolliffe (2012) "Identifying the Effects of SNAP (Food Stamps) on Child Health Outcomes When Participation Is Endogenous and Misreported," *Journal of the American Statistical Association,* 107:499, 958-975.

Lewbel, Arthur (2007) "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75(2):537-551.

Mahajan, Aprajit (2006) "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74(3):631-665.

Mellow, Wesley and Hal Sider (1983) "Accuracy of Response in Labor Market Surveys: Evidence and Implications," *Journal of Labor Economics*, 1(4):331-344.

Newey, Whitney K. (1984) "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters*, 14:201-6.

Nguimkeu, Pierre, Augustine Denteh, and Rusty Tchernis (2019) "On the Estimation of Treatment Effects with Endogenous Misreporting," *Journal of Econometrics,* 208:487-506.

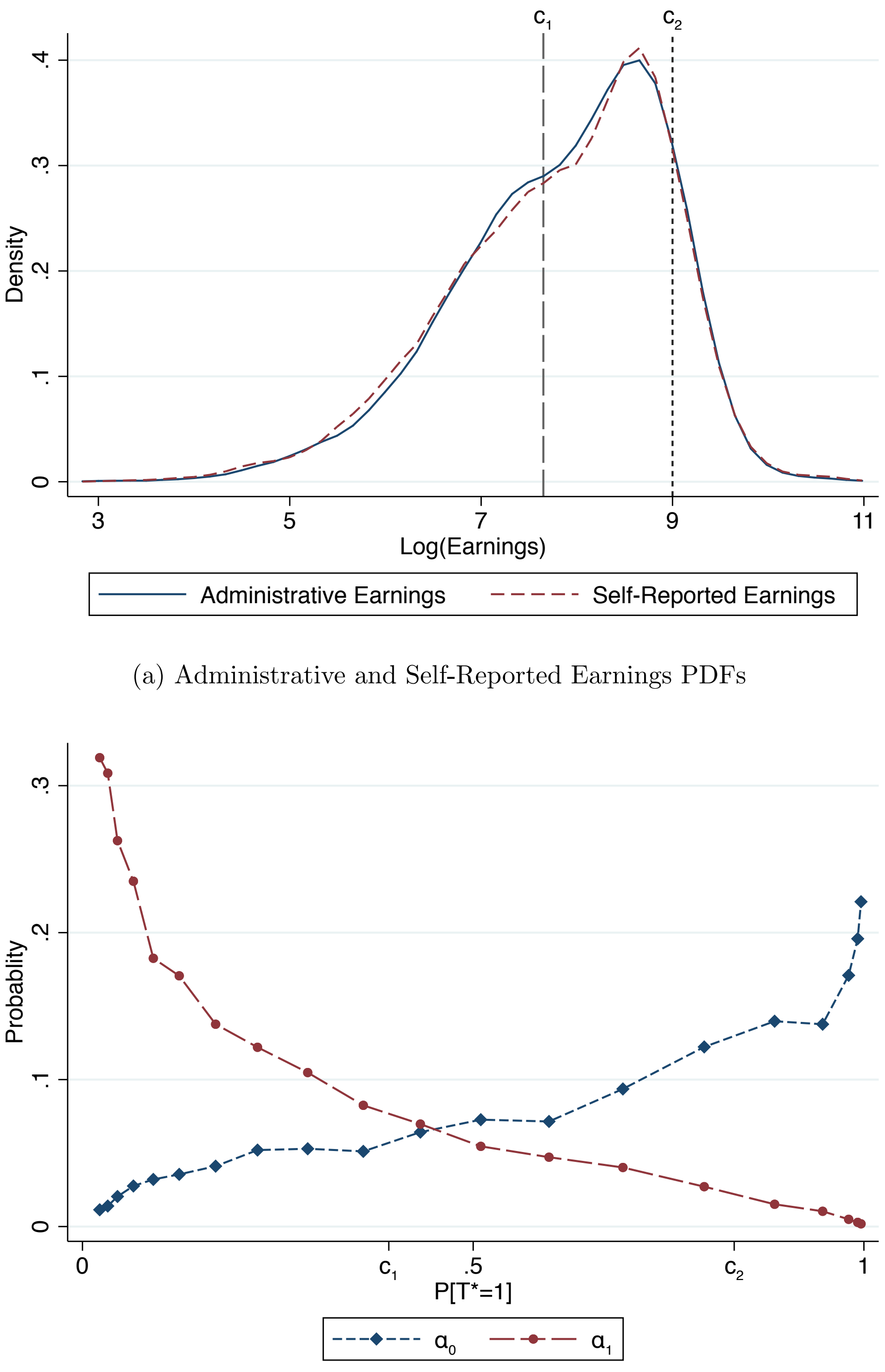
Olsen, Randall J. (1980) "A Least Squares Correction for Selectivity Bias," *Econometrica*, 48(7): 1815-1820.

Social Security Administration. Current Population Survey, 1973, and Social Security Records: Exact Match Data. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2005-11-04.

Ura, Takuya (2018) "Heterogeneous Treatment Effects with Mismeasured Endogenous Treatment," *Quantitative Economics*, 9(3):1335-1370.

Yanagi, Takahide (2019) "Inference on Local Average Treatment Effects for Misclassified Treatment," *Econometric Reviews*, 38(8):938-960.

Figure 1: Misclassification Example - 1973 CPS-IRS Matched Earnings

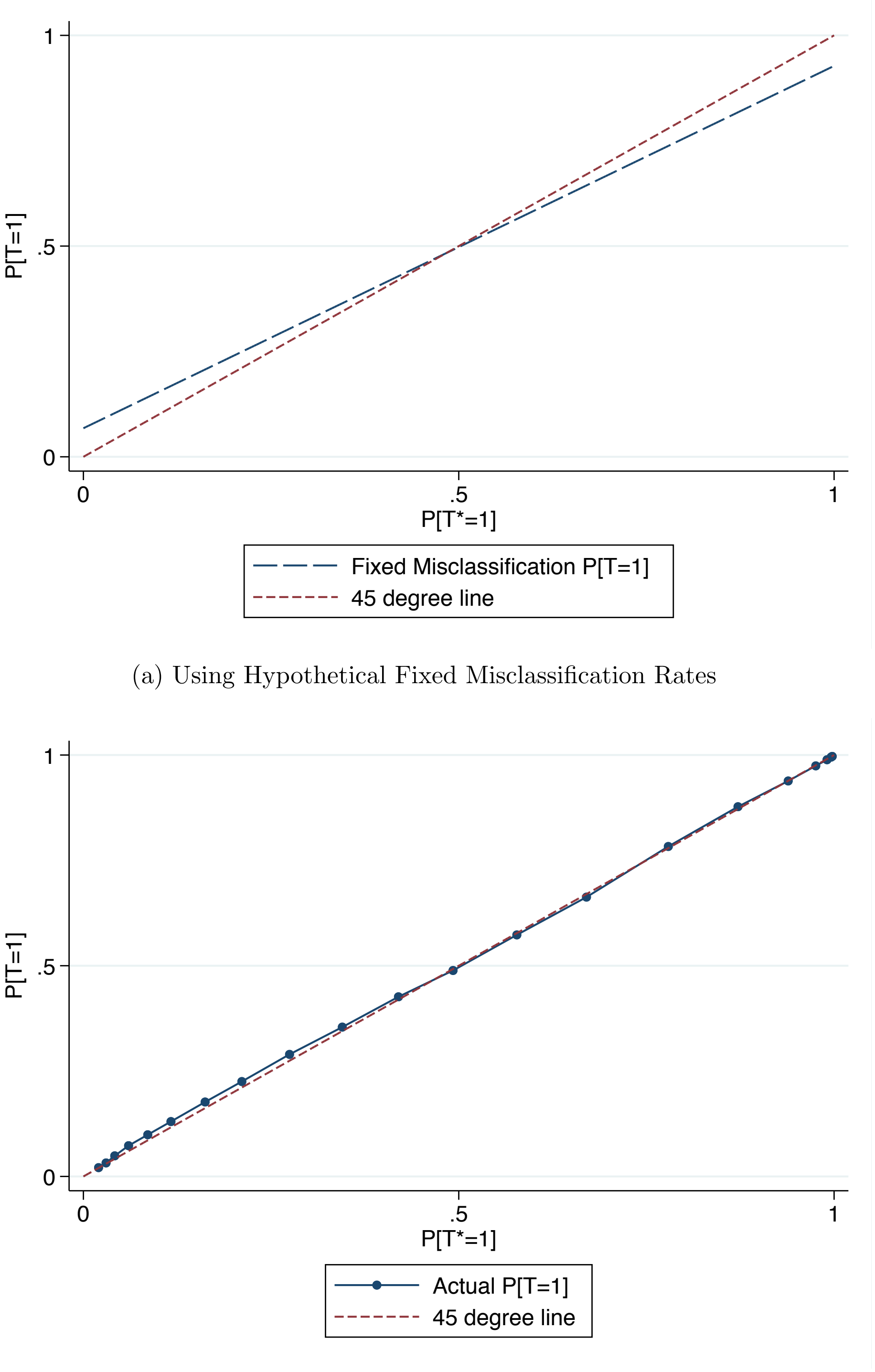


(a) Administrative and Self-Reported Earnings PDFs

(b) Misclassification Rates

Notes: Both panels use data from the 1973 CPS-IRS matched earnings data. Details about the data are discussed in the text and in Appendix section A.1. Panel 1a plots the PDFs of both self-reported earnings and earnings reported on IRS tax returns from a sample of single men and women. Panel 1b reports misclassification rates; the details for constructing these rates can be found in Section 2.

Figure 2: Misclassification Example - 1973 CPS-IRS Matched Earnings - $P[T = 1]$ vs. $P[T^* = 1]$

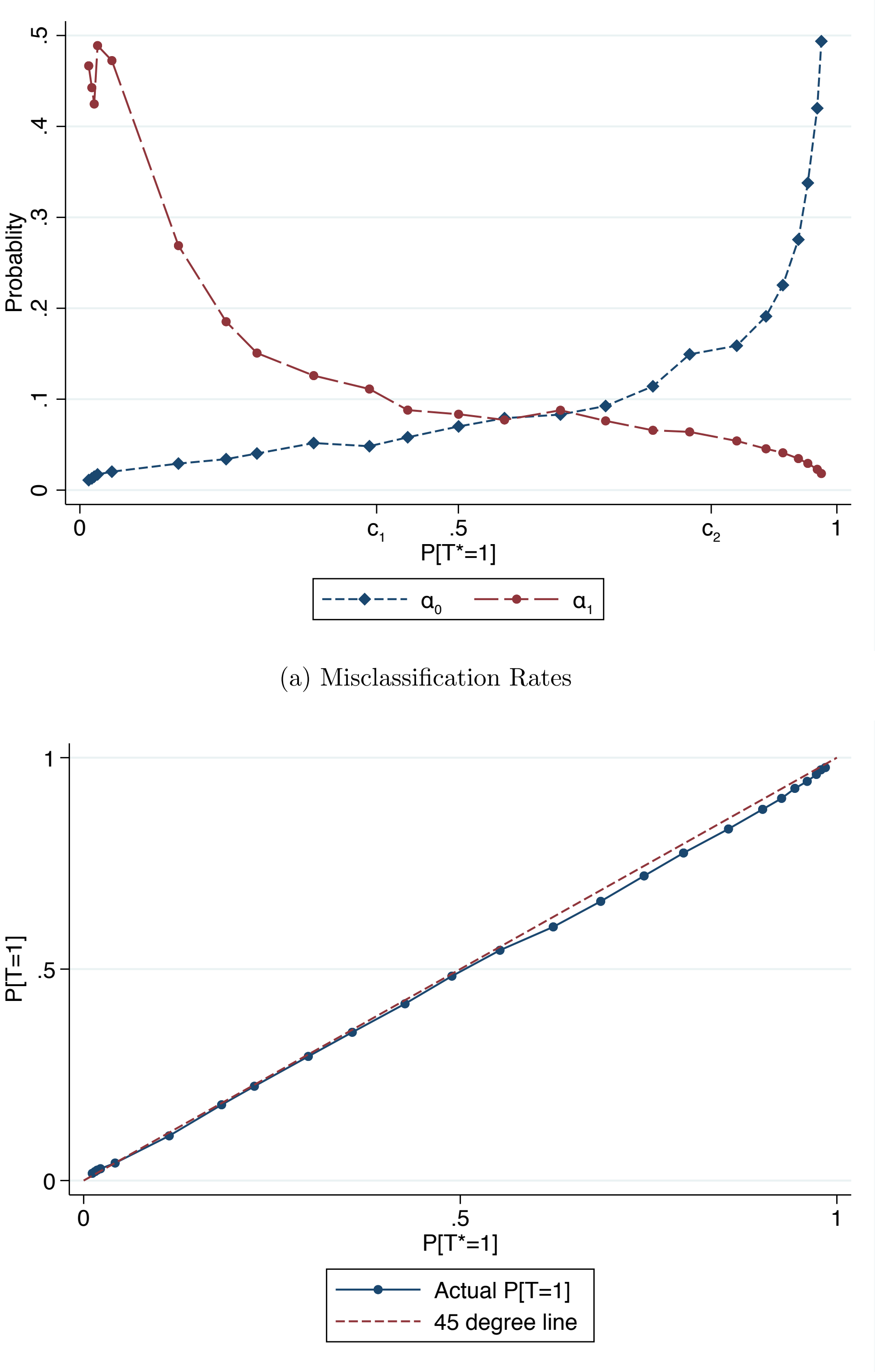


(a) Using Hypothetical Fixed Misclassification Rates

(b) Using Actual Misclassification Rates

Notes: Panel 2a plots the hypothetical relationship between $P[T = 1]$ and $P[T^* = 1]$ under the assumption of fixed misclassification rates, where the α_0 and α_1 used to construct this figure correspond to the threshold c_1 shown in Figure 1. Panel 2b plots the actual relationship between $P[T = 1]$ and $P[T^* = 1]$ found in the 1973 CPS-IRS matched earnings data. Details about the data are discussed in the text and in Appendix section A.1.

Figure 3: Misclassification Example - 1977 January CPS Matched Employer-Employee Wages

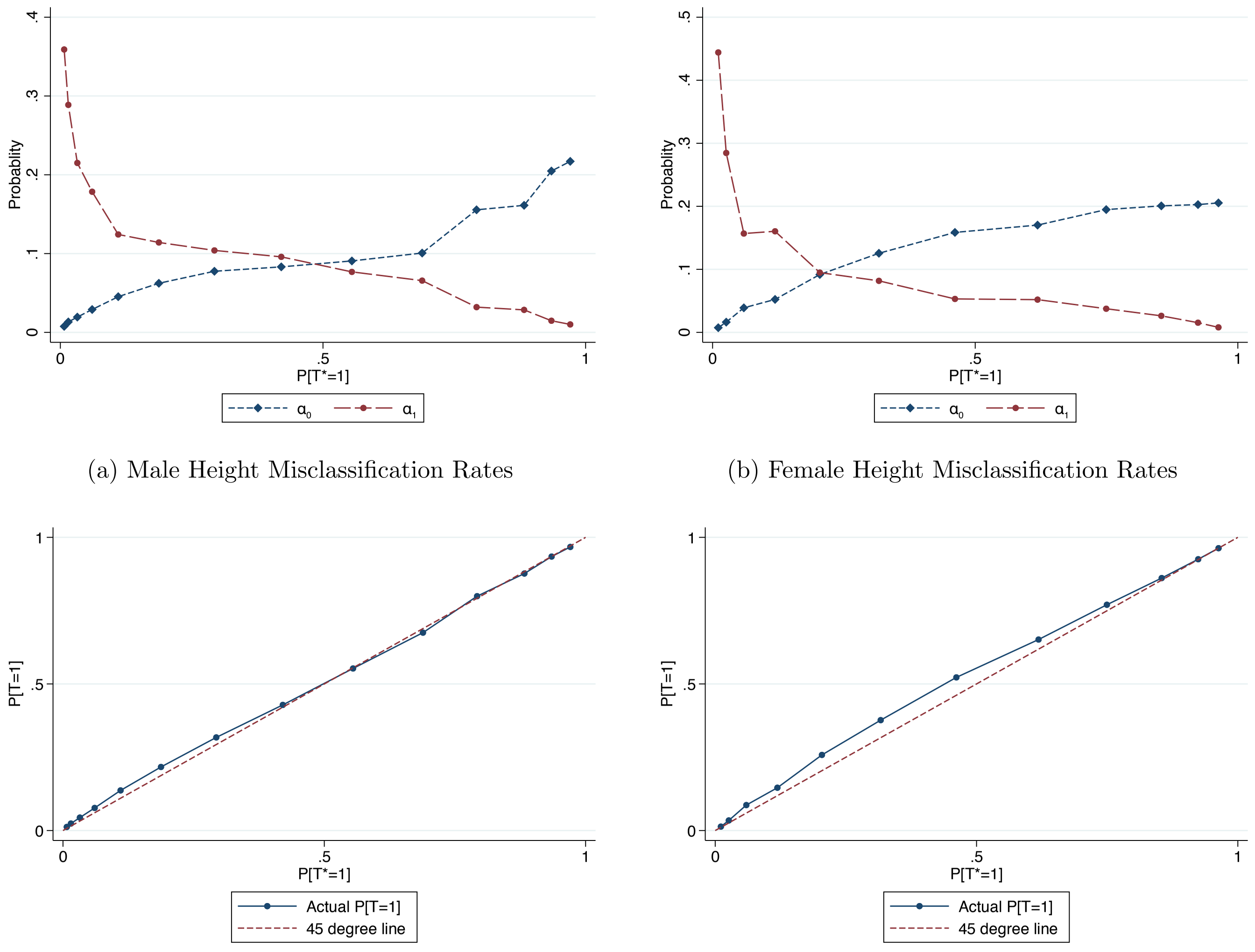


(a) Misclassification Rates

(b) $P[T = 1]$ vs. $P[T^* = 1]$

Notes: Both panels use data from the January 1977 CPS matched employer-employee wage data. Details about the data are discussed in the text and in Appendix section A.1. Panel 3a reports misclassification rates, with the details for constructing these rates found in Section 2. Panel 3b plots the relationship between $P[T = 1]$ and $P[T^* = 1]$.

Figure 4: Misclassification Example - Continuous NHANES Height



(a) Male Height Misclassification Rates

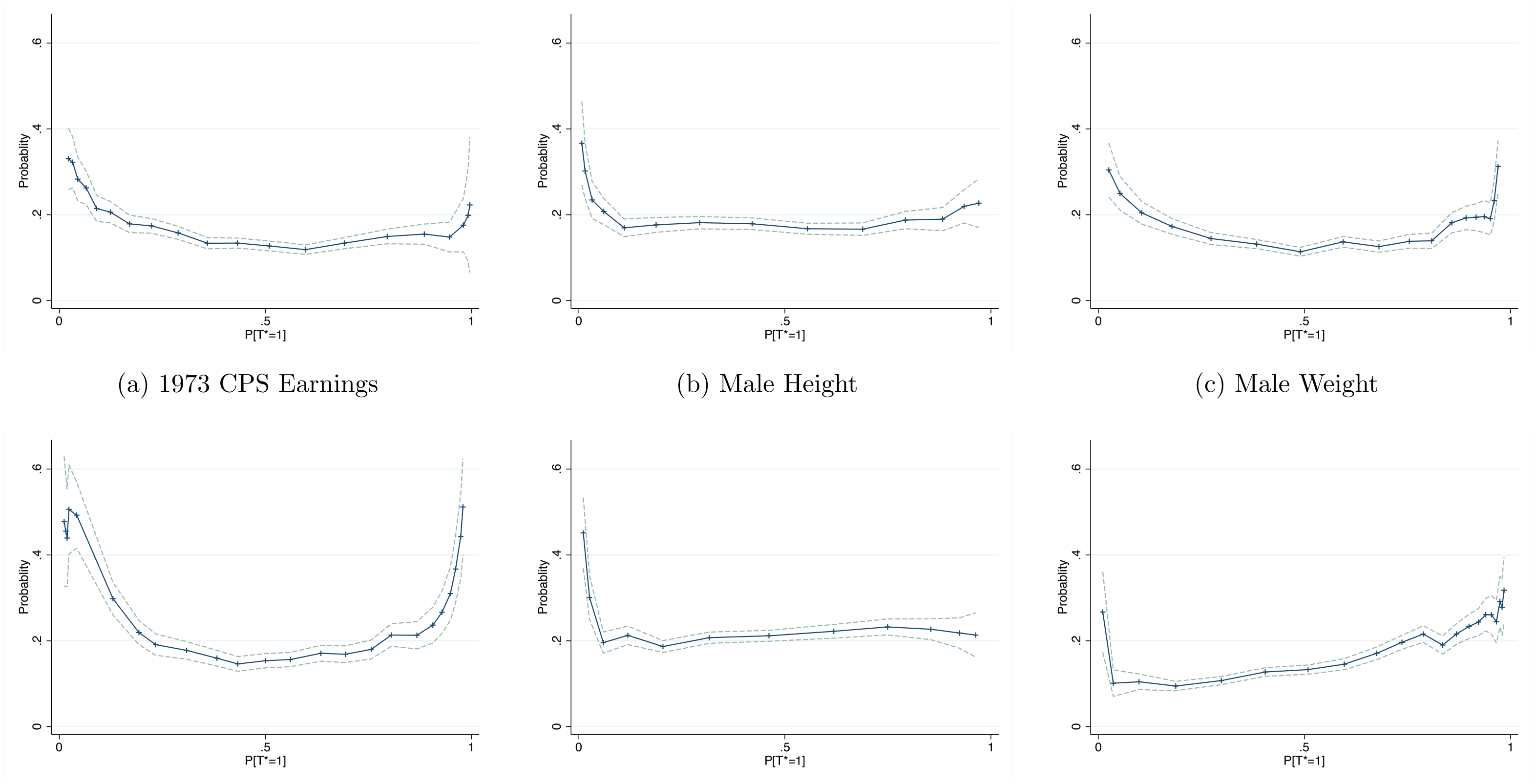
(b) Female Height Misclassification Rates

(c) Male Height $P[T = 1]$ vs. $P[T^* = 1]$

(d) Female Height $P[T = 1]$ vs. $P[T^* = 1]$

Notes: The panels use both self-reported and measured height from Continuous NHANES data. Details about the data are discussed in the text and in Appendix section A.1. Panels 4a and 4b report misclassification rates for men and women, respectively, with the details for constructing these rates found in Section 2. Panels 4c and 4d plots the relationship between $P[T = 1]$ and $P[T^* = 1]$ for men and women, respectively.

Figure 5: Misclassification Example - Sum of the Misclassification Rates



(a) 1973 CPS Earnings

(b) Male Height

(c) Male Weight

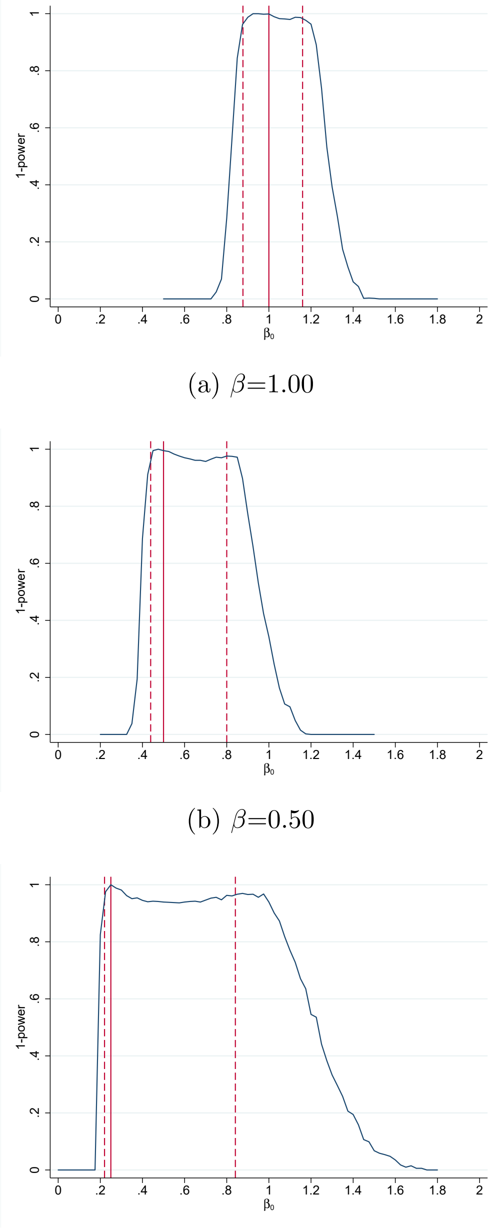
(d) 1977 CPS Wages

(e) Female Height

(f) Female Weight

Notes: The solid lines in the panels present the sum of the misclassification rates. The dashed lines in each panel are the bootstrapped 95% confidence intervals.

Figure 6: Monte Carlo Simulations - Coverage Curves for “Small Samples”



(a) β=1.00

(b) β=0.50

(c) β=0.25

Notes: The coverage curves are constructed based on 1,000 iterations for $N = 1,000$. The vertical solid line in each panel is the true β used to generate the data. The dashed vertical lines are the lower and upper bounds on β, which are incorporated as inequality moments.

Table 1: Monte Carlo Simulations - Varying N and β

N	0.0	0.25	0.5	1.0	1.5	2.0
Panel A: Proportion GMM Converges						
1,000	0.970	0.844	0.885	0.940	0.969	0.978
10,000	0.969	0.979	0.999	1.000	1.000	1.000
100,000	0.951	1.000	1.000	1.000	1.000	1.000
Panel B: Mean of β if Converges						
1,000	-0.009	0.755	0.984	1.427	2.021	2.567
10,000	0.048	0.385	0.539	1.017	1.517	2.018
100,000	0.006	0.256	0.504	1.003	1.502	2.001
Panel C: Mean of SE of β if Converges						
1,000	1350	636.8	1346	15.04	15.06	11.71
10,000	1170	33.50	0.300	0.113	0.137	0.168
100,000	495.7	0.026	0.026	0.032	0.040	0.049
Panel D: Coverage of True β if Converges						
1,000	0.992	0.906	0.884	0.911	0.918	0.921
10,000	0.995	0.892	0.925	0.925	0.943	0.946
100,000	1.000	0.948	0.954	0.954	0.951	0.952

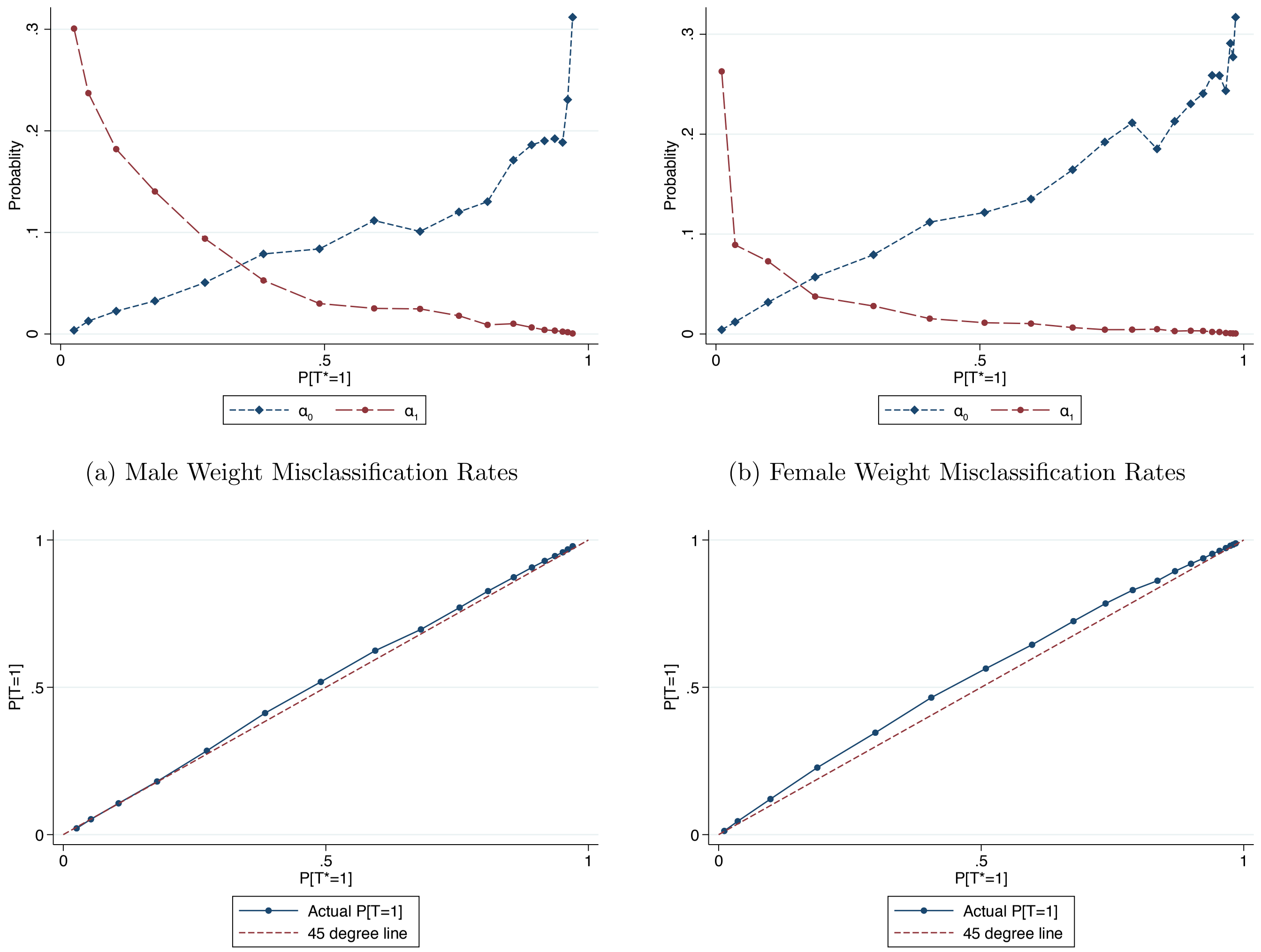
Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (14). Each simulation uses 1,000 iterations. The numbers within each of the four panels correspond to a different simulation, where the sample size N and the parameter of interest β vary across simulations. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges. Panel D reports the coverage for β across the iterations where the estimator converges.

Table 2: Monte Carlo Simulations - Varying α's

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Structure of data generating process						
α_{00}	0.070	0.055	0.035	0.059	0.023	0.014
α_{01}	0.070	0.070	0.070	0.078	0.046	0.039
α_{02}	0.070	0.085	0.105	0.097	0.065	0.059
$\alpha_{00} + \alpha_{10}$	0.130	0.130	0.130	0.156	0.165	0.173
$\alpha_{01} + \alpha_{11}$	0.130	0.130	0.130	0.155	0.158	0.159
$\alpha_{02} + \alpha_{12}$	0.130	0.130	0.130	0.156	0.156	0.156
Panel B: Estimates of β						
OLS	0.870	0.877	0.886	0.853	0.844	0.842
	(.002)	(.002)	(.002)	(.002)	(.003)	(.003)
2SLS	1.150	1.032	0.907	1.030	1.014	1.004
	(.011)	(.009)	(.008)	(.009)	(.008)	(.008)
Our estimator	1.003	1.003	1.002	0.998	1.006	1.012
	(.033)	(.033)	(.033)	(.033)	(.046)	(.052)
Our lower bound	0.928	0.913	0.897	0.895	0.888	0.884
	(.003)	(.003)	(.003)	(.003)	(.003)	(.004)

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (14). Each simulation is based on 1,000 iterations, with a sample size of 100,000 observations. The mis-measured variables T are constructed so that misclassification rate α_{0j} and the sum of the misclassification rates $\alpha_{0j} + \alpha_{1j}$ for instrument value j match the values shown in the top panel. The values for the remaining parameters are discussed in Section 5. Panel B contains average estimates of β across iterations for the listed estimation methods. The standard deviation of the estimates across iterations are shown in parentheses.

Figure A1: Misclassification Example - Continuous NHANES Weight



(a) Male Weight Misclassification Rates

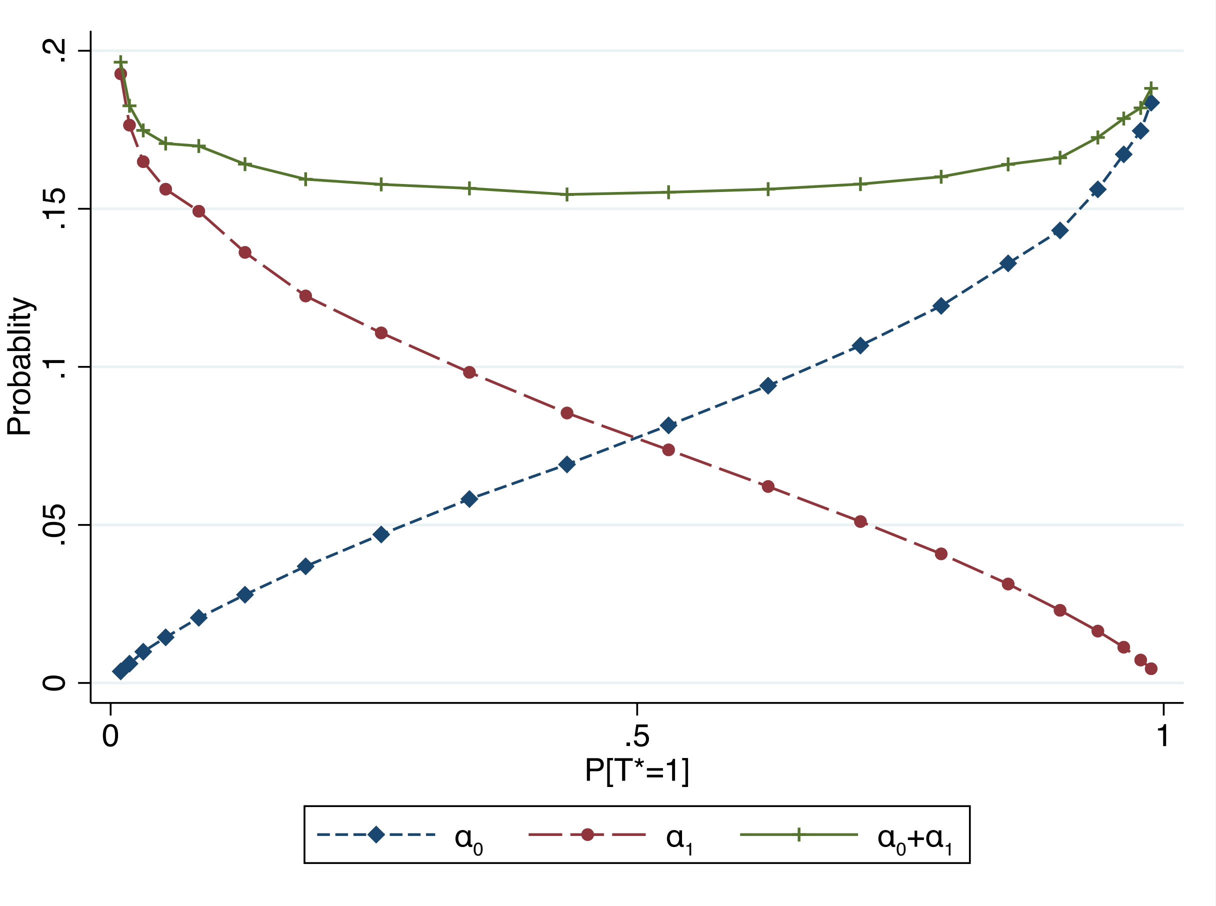
(b) Female Weight Misclassification Rates

(c) Male Weight $P[T = 1]$ vs. $P[T^* = 1]$

(d) Female Weight $P[T = 1]$ vs. $P[T^* = 1]$

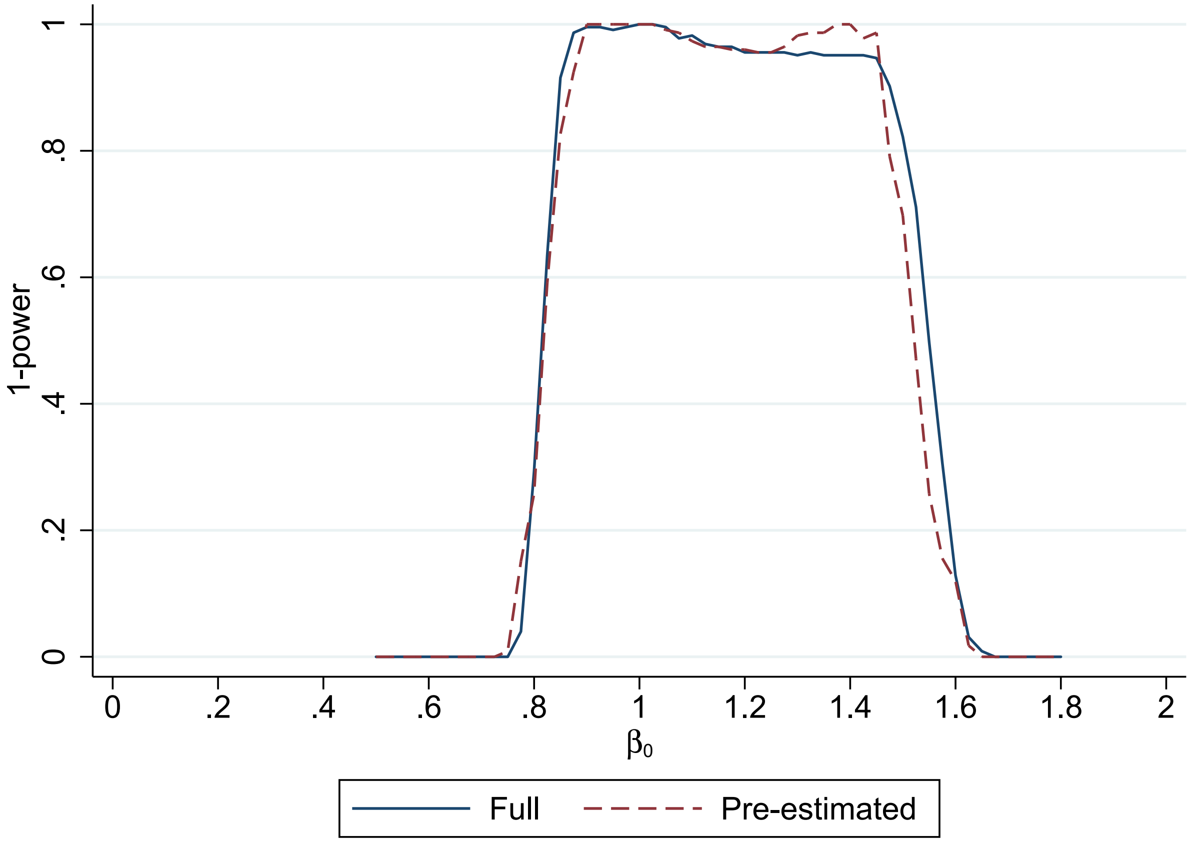
Notes: The panels use both self-reported and measured weight from Continuous NHANES data. Details about the data are discussed in the text and in Appendix section A.1. Panels A1a and A1b report misclassification rates for men and women, respectively, with the details for constructing these rates found in Section 2. Panels A1c and A1d plots the relationship between $P[T = 1]$ and $P[T^* = 1]$ for men and women, respectively.

Figure A2: Simulated Misclassification Rates Using a Mis-Measured Index



Notes: To construct this figure, we generate the mis-measured index E by adding a mean zero measurement error ν to the random variable E^*. For any given threshold c, we construct the binary indicators $T^* = 1(E^* \leq c)$ and $T = 1(E \leq c)$ and the corresponding misclassification rates as discussed in the text. This figure shows these values average over 500,000 iterations.

Figure A3: Coverage Curves with Full and Pre-Estimated Moments



Notes: This figure is based on 1,000 iterations for $N = 1,000$. The lower bound is specified to be the OLS regression of Y on T and the upper bound is specified to be the inverse of the reverse OLS regression of T on Y. The coverage curve marked "Full" does not pre-estimate the bounds, whereas the coverage curve marked "Pre-estimated" pre-estimates the bounds and then inflates the variance of the moment to account for pre-estimation through a bootstrapped procedure.

Table A1: Monte Carlo Simulations - Varying α and β, $N = 1,000$

$\alpha_{0j} + \alpha_{0j}$	β=0.0	β=0.25	β=0.5	β=1.0	β=1.5	β=2.0
Panel A: Proportion GMM Converges						
0.00	0.964	0.847	0.907	0.988	0.997	1.000
0.13a	0.970	0.844	0.885	0.940	0.969	0.978
0.13b	0.964	0.838	0.897	0.944	0.971	0.967
0.26a	0.963	0.825	0.863	0.916	0.925	0.930
0.26b	0.968	0.848	0.869	0.914	0.926	0.937
0.39a	0.960	0.827	0.833	0.865	0.899	0.914
0.39b	0.974	0.828	0.843	0.891	0.895	0.905
Panel B: Mean of β if Converges						
0.00	0.003	0.679	0.883	1.187	1.586	2.064
0.13a	-0.010	0.755	0.984	1.427	2.021	2.567
0.13b	-0.001	0.896	1.000	1.496	2.098	2.559
0.26a	0.013	0.825	1.131	1.812	2.385	2.990
0.26b	-0.020	0.917	1.135	1.732	2.412	2.941
0.39a	-0.060	0.779	1.288	1.911	2.689	3.337
0.39b	-0.020	0.886	1.128	1.915	2.666	3.385
Panel C: Mean of SE of β if Converges						
0.00	6347	252.8	212.1	3.158	0.480	0.425
0.13a	1350	636.8	1346	15.04	15.06	11.71
0.13b	3011	3305	97.77	21.78	20.66	15.54
0.26a	3914	1017	412.7	232.7	66.74	46.41
0.26b	590.2	2499	1557	77.15	51.98	47.38
0.39a	2892	343.3	1362	183.3	207.7	87.83
0.39b	3655	5147	281.4	92.79	129.7	103.2

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (14). Each simulation uses 1,000 iterations. The numbers within each of the three panels correspond to a different simulation, varying the parameters β and $\alpha_{0j} + \alpha_{1j}$ and keeping the sample size fixed at $N = 1,000$. Each $\alpha_{0j} + \alpha_{1j}$ value has two specifications for α_{0j}, with α_{1j} then being set to equal the designated sum. The α_{0j} values for 0.13a are (0.055, 0.070, 0.085) and for 0.13b are (0.035, 0.070, 0.105), following columns (2) and (3) of Table 2; the values for 0.26a and 0.26b multiply these values by 2, respectively, and the values for 0.39a and 0.39b multiply these values by 3, respectively. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges.

Table A2: Monte Carlo Simulations - Varying α and β, $N = 10,000$

$\alpha_{0j} + \alpha_{0j}$	β=0.0	β=0.25	β=0.5	β=1.0	β=1.5	β=2.0
Panel A: Proportion GMM Converges						
0.00	0.963	0.985	1.000	1.000	1.000	1.000
0.13a	0.960	0.979	0.999	1.000	1.000	1.000
0.13b	0.965	0.973	1.000	1.000	1.000	1.000
0.26a	0.969	0.952	0.997	0.999	1.000	1.000
0.26b	0.971	0.956	0.992	0.999	0.999	1.000
0.39a	0.968	0.928	0.986	0.995	0.995	0.998
0.39b	0.975	0.937	0.982	0.994	0.996	0.995
Panel B: Mean of β if Converges						
0.00	0.005	0.330	0.515	1.006	1.504	2.003
0.13a	0.048	0.385	0.539	1.017	1.517	2.018
0.13b	-0.010	0.357	0.537	1.017	1.516	2.017
0.26a	-0.001	0.414	0.566	1.042	1.571	2.059
0.26b	-0.001	0.412	0.551	1.039	1.553	2.062
0.39a	-0.001	0.432	0.653	1.130	1.647	2.217
0.39b	-0.001	0.437	0.624	1.109	1.672	2.208
Panel C: Mean of SE of β if Converges						
0.00	2962	20.06	0.082	0.066	0.065	0.065
0.13a	1170	33.50	0.300	0.113	0.137	0.168
0.13b	1045	14.46	0.24	0.113	0.138	0.168
0.26a	518.7	58.63	0.687	0.205	1.168	0.302
0.26b	649.6	48.01	0.244	0.186	0.264	0.315
0.39a	842.6	14.38	9.299	1.414	0.786	1.618
0.39b	329.6	13.57	3.114	0.763	1.730	0.985

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (14). Each simulation uses 1,000 iterations. The numbers within each of the three panels correspond to a different simulation, varying the parameters β and $\alpha_{0j} + \alpha_{1j}$ and keeping the sample size fixed at $N = 10,000$. Each $\alpha_{0j} + \alpha_{1j}$ value has two specifications for α_{0j}, with α_{1j} then being set to equal the designated sum. The α_{0j} values for 0.13a are (0.055, 0.070, 0.085) and for 0.13b are (0.035, 0.070, 0.105), following columns (2) and (3) of Table 2; the values for 0.26a and 0.26b multiply these values by 2, respectively, and the values for 0.39a and 0.39b multiply these values by 3, respectively. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges.

Table A3: Monte Carlo Simulations - Varying α and β, $N = 100,000$

$\alpha_{0j} + \alpha_{0j}$	β=0.0	β=0.25	β=0.5	β=1.0	β=1.5	β=2.0
Panel A: Proportion GMM Converges						
0.00	0.963	1.000	1.000	1.000	1.000	1.000
0.13a	0.951	1.000	1.000	1.000	1.000	1.000
0.13b	0.976	1.000	1.000	1.000	1.000	1.000
0.26a	0.959	1.000	1.000	1.000	1.000	1.000
0.26b	0.974	1.000	1.000	1.000	1.000	1.000
0.39a	0.953	1.000	1.000	0.999	1.000	1.000
0.39b	0.965	1.000	1.000	1.000	1.000	1.000
Panel B: Mean of β if Converges						
0.00	0.007	0.254	0.503	1.002	1.502	2.002
0.13a	0.006	0.256	0.504	1.003	1.502	2.001
0.13b	0.003	0.256	0.504	1.002	1.502	2.001
0.26a	0.001	0.258	0.505	1.004	1.504	2.003
0.26b	0.020	0.258	0.505	1.005	1.504	2.004
0.39a	0.003	0.268	0.509	1.009	1.510	2.027
0.39b	-0.001	0.265	0.509	1.008	1.511	2.016
Panel C: Mean of SE of β if Converges						
0.00	255	0.022	0.020	0.020	0.020	0.020
0.13a	496	0.026	0.026	0.032	0.040	0.049
0.13b	337	0.026	0.026	0.032	0.040	0.050
0.26a	552	0.033	0.033	0.045	0.059	0.075
0.26b	479	0.033	0.033	0.045	0.060	0.076
0.39a	492	0.178	0.044	0.062	0.084	0.121
0.39b	209	0.050	0.044	0.061	0.085	0.110

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (14). Each simulation uses 1,000 iterations. The numbers within each of the three panels correspond to a different simulation, varying the parameters β and $\alpha_{0j} + \alpha_{1j}$ and keeping the sample size fixed at $N = 100,000$. Each $\alpha_{0j} + \alpha_{1j}$ value has two specifications for α_{0j}, with α_{1j} then being set to equal the designated sum. The α_{0j} values for 0.13a are (0.055, 0.070, 0.085) and for 0.13b are (0.035, 0.070, 0.105), following columns (2) and (3) of Table 2; the values for 0.26a and 0.26b multiply these values by 2, respectively, and the values for 0.39a and 0.39b multiply these values by 3, respectively. The values for the remaining parameters are discussed in Section 5. Panel A reports the proportion of times that the GMM estimator converges across iterations. Panel B reports the average of β across the iterations where the estimator converges. Panel C reports the average standard error of β across the iterations where the estimator converges.

Table A4: Comparing Proposed Estimators for Mis-Classified Binary Variables

	(1)	(2)	(3)	(4)	(5)
Panel A: Structure of data generating process					
α_{00}	0.070	0.055	0.055	0.055	0.055
α_{01}	0.070	0.070	0.070	0.070	0.070
α_{02}	0.070	0.085	0.085	0.085	0.085
$\alpha_{00} + \alpha_{10}$	0.130	0.130	0.130	0.130	0.130
$\alpha_{01} + \alpha_{11}$	0.130	0.130	0.130	0.130	0.130
$\alpha_{02} + \alpha_{12}$	0.130	0.130	0.130	0.130	0.130
Homoskedastic errors?	Yes	Yes	No	Yes	No
Symmetric errors?	Yes	Yes	Yes	No	No
Panel B: Estimates of β					
Our estimator	1.003	1.003	1.003	1.002	1.000
	(.033)	(.033)	(.051)	(.042)	(.034)
Frazis and Lowenstein (2003)	1.000	0.951	0.951	0.951	0.951
	(.005)	(.005)	(.007)	(.006)	(.004)
Mahajan (2006)	1.000	0.951	0.951	0.951	0.951
	(.005)	(.005)	(.007)	(.006)	(.005)
Lewbel (2007)	0.999	0.896	0.897	0.896	0.899
	(.038)	(.034)	(.049)	(.041)	(.035)
Chen, Hu, and Lewbel (2008a)	1.000	1.000	1.150	1.000	0.964
	(.002)	(.002)	(.004)	(.003)	(.002)
Chen, Hu, and Lewbel (2008b)	1.000	1.000	1.027	0.981	0.854
	(.002)	(.002)	(.014)	(.003)	(.009)
DiTraglia and Garcia-Jimeno (2019)	1.001	1.001	1.150	1.000	0.965
	(.009)	(.008)	(.016)	(.013)	(.011)

Notes: This table reports Monte Carlo simulations from using GMM to estimate the moment conditions found in (14). Each simulation is based on 1,000 iterations, with a sample size of 100,000 observations. The mis-measured variables T are constructed so that misclassification rate α_{0j} and the sum of the misclassification rates $\alpha_{0j} + \alpha_{1j}$ for instrument value j match the values shown in the top panel. The values for the remaining parameters are discussed in Section 5. See Appendix section A.8 for a complete description of the data generating processes specified for each column. Panel B contains average estimates of β across iterations for the listed estimation method. The standard deviation of the estimates across iterations are shown in parentheses.

A Appendix

A.1 Dataset Descriptions

We use the March 1973 CPS that is linked both to Social Security earnings records and to tax returns filed with the IRS. The earnings data used in our examples refer to the prior calendar year, 1972. The available SS earnings records are top-coded at the annual ceiling for earnings subject to SS tax, which is $9,000 in 1972. Nearly half of privately employed men in the sample have top-coded SS earnings data. On the other hand, the IRS earnings records and self-reported earnings in the CPS are top-coded at $50,000, which affects less than one-half of one percent of privately employed men in the data.

A drawback to the available IRS earnings data is that these are taken directly from tax filings, which only have a single item in which the combined earnings of all household members is reported. To circumvent this issue, we restrict our sample to men and women who designate their household status as single on the tax return such that the IRS earnings amount should only reflect their own earnings. In addition, we restrict the analysis to men and women who are report to the CPS that they are privately employed, have positive IRS and CPS earnings, and have non-imputed CPS earnings. We also restrict the data to individuals in households that are deemed to be "good" matches (where the variable V1255 is not equal to 4) and to individuals with matches to available IRS data (the variable V1253 equals 0). Our final sample includes 8,031 single men and women. All our analyses are weighted using the "final" CPS-IRS-SSA STATS unit administrative weight (the variable V1264), which account for selection into the CPS and the match between the CPS and the administrative records.

We also use data from a special supplement to the January 1977 CPS. A subsample of survey respondents was asked additional questions regarding union status, earnings, and hours worked. These individuals were also asked to provide the name and address of their employer. Their employers were subsequently asked to provide information for the survey respondent, including the aforementioned variables. For a more detailed discussion of this data, see Mellow and Sider (1983). Using code graciously provided to us by Alan Krueger, we employ the same sample restrictions as Angrist and Krueger (1999).

We also use the Continuous National Health and Nutrition Examination Survey, which is a survey of the health and diet of Americans conducted by the National Center for Health Statistics which started in 1999 and is conducted on a two-year cycle. This survey replaced the prior National Health and Nutrition Examination Surveys which had previously been fielded on an idiosyncratic schedule. Households and sample persons are selected using a stratified, multistage sampling design for each cycle. A screener and basic questionnaire are completed as part of initial in-home visits. A randomizing computer algorithm selects a sample person from each household roster. Survey respondents complete a questionnaire covering demographic, dietary, socioeconomic, and health topics. During this in-home interview, respondents are also asked to provide their height and weight. Following the in-home interview, the selected sample persons make appointments to visit a Mobile Examination Center for a detailed physical examination. During this subsequent physical examination, the individual's height and weight are measured by survey staff. All of our analyses are weighted to account for selection into the NHANES and completing both the in-home interview and the physical exam (the variable WTMEC2YR).

A.2 Derivation of Equations (5) and (6)

For a given value of Z, the expected value of the outcome when $T = 1$ is

$$\begin{aligned}
E\left[Y|T=1,Z=j\right] &= E\left[Y|T^*=0,T=1,Z=j\right]\cdot P[T^*=0|T=1,Z=j] \\
&+ E\left[Y|T^*=1,T=1,Z=j\right]\cdot P[T^*=1|T=1,Z=j] \\
&= E\left[Y|T^*=0,Z=j\right]\cdot P[T^*=0|T=1,Z=j] \\
&+ E\left[Y|T^*=1,Z=j\right]\cdot P[T^*=1|T=1,Z=j] \\
&= E\left[Y|T^*=0,Z=j\right]\cdot \frac{P[T=1|T^*=0,Z=j]\cdot P[T^*=0|Z=j]}{P[T=1|Z=j]} \\
&+ E\left[Y|T^*=1,Z=j\right]\cdot \frac{P[T=1|T^*=1,Z=j]\cdot P[T^*=1|Z=j]}{P[T=1|Z=j]} \\
&= E\left[Y|T^*=0,Z=j\right]\cdot \frac{\alpha_{0j}\left(1-p_j^*\right)}{p_j} + E\left[Y|T^*=1,Z=j\right]\cdot \frac{\left(1-\alpha_{1j}\right)p_j^*}{p_j}
\end{aligned} \tag{19}$$

The second equality follows from the non-differential measurement error assumption and the third equality follows from Bayes Theorem. Similarly,

$$E\left[Y|T=0,Z=j\right]=E\left[Y|T^{*}=0,Z=j\right]\cdot\frac{\left(1-\alpha_{0j}\right)\left(1-p_{j}^{*}\right)}{1-p_{j}}+E\left[Y|T^{*}=1,Z=j\right]\cdot\frac{\alpha_{1j}\cdot p_{j}^{*}}{1-p_{j}} \quad (20)$$

We can solve for $E\left[Y|T^{*}=1,Z=j\right]$ and $E\left[Y|T^{*}=0,Z=j\right]$ using equations (19) and (20) in terms of both observed quantities ($E\left[Y|T=0,Z=j\right]$, $E\left[Y|T=1,Z=j\right]$, and p_j) and unobserved quantities (α_{0j}, α_{1j}, and p_j^*)

$$\begin{aligned}E\left[Y|T^{*}=1,Z=j\right]&=\frac{p_{j}E\left[Y|T=1,Z=j\right]-\alpha_{0j}E\left[Y|Z=j\right]}{p_{j}^{*}\left(1-\alpha_{0j}-\alpha_{1j}\right)}\\E\left[Y|T^{*}=0,Z=j\right]&=\frac{\left(1-p_{j}\right)E\left[Y|T=0,Z=j\right]-\alpha_{1j}E\left[Y|Z=j\right]}{\left(1-p_{j}^{*}\right)\left(1-\alpha_{0j}-\alpha_{1j}\right)}\end{aligned}$$

which use the substitution $E\left[Y|Z=j\right]=E\left[Y|T=1,Z=j\right]p_{j}+E\left[Y|T=0,Z=j\right]\left(1-p_{j}\right)$.

Noting that $p_{j}-\alpha_{0j}=\left(1-\alpha_{0j}-\alpha_{1j}\right)p_{j}^{*}$, which follows immediately from the equation for p_j, we can simplify the expression for $E\left[Y|T^{*}=1,Z=j\right]$ to yield

$$\begin{aligned}E\left[Y|T^{*}=1,Z=j\right]&=\frac{p_{j}E\left[Y|T=1,Z=j\right]-\alpha_{0j}E\left[Y|Z=j\right]}{p_{j}^{*}\left(1-\alpha_{0j}-\alpha_{1j}\right)}\\&=\frac{p_{j}E\left[Y|T=1,Z=j\right]-\alpha_{0j}E\left[Y|Z=j\right]}{p_{j}-\alpha_{0j}}\\&=\frac{p_{j}E\left[Y|T=1,Z=j\right]-\alpha_{0j}E\left[Y|Z=j\right]+\left(p_{j}E\left[Y|Z=j\right]-p_{j}E\left[Y|Z=j\right]\right)}{p_{j}-\alpha_{0j}}\\&=\frac{p_{j}E\left[Y|T=1,Z=j\right]-p_{j}E\left[Y|Z=j\right]+\left(p_{j}-\alpha_{0j}\right)E\left[Y|Z=j\right]}{p_{j}-\alpha_{0j}}\\&=\frac{E\left[YT|Z=j\right]-p_{j}E\left[Y|Z=j\right]}{p_{j}-\alpha_{0j}}+E\left[Y|Z=j\right]\\&=\frac{COV\left(Y,T|Z=j\right)}{p_{j}-\alpha_{0j}}+E\left[Y|Z=j\right] \quad (21)\end{aligned}$$

where $COV\left(Y,T|Z=j\right)$ is the covariance of Y and T among those observations where $Z=j$ and making of the fact that

$$E\left[YT|Z=j\right]=p_{j}E\left[Y\cdot 1|T=1,Z=j\right]+\left(1-p_{j}\right)E\left[Y\cdot 0|T=0,Z=j\right]=p_{j}E\left[Y|T=1,Z=j\right]$$

The following expression follows similarly

$$E\left[Y|T^{*}=0,Z=j\right]=E\left[Y|Z=j\right]-\frac{COV\left(Y,T|Z=j\right)}{1-p_{j}-\alpha_{1j}} \tag{22}$$

A.3 Derivation of Equation (7)

Subtracting equation (4) from equation (3) and then inserting equations (5) and (6) into the resulting expression yields

$$\begin{gathered}
E\left[Y|T^{*}=1,Z=j\right]-E\left[Y|T^{*}=0,Z=j\right]=E\left[Y|T^{*}=1,Z=k\right]-E\left[Y|T^{*}=0,Z=k\right]\\
\frac{COV\left(Y,T|Z=j\right)}{p_{j}-\alpha_{0j}}+\frac{COV\left(Y,T|Z=j\right)}{1-p_{j}-\alpha_{1j}}=\frac{COV\left(Y,T|Z=k\right)}{p_{k}-\alpha_{0k}}+\frac{COV\left(Y,T|Z=k\right)}{1-p_{k}-\alpha_{1k}}\\
\frac{COV\left(Y,T|Z=j\right)}{\left(1-\alpha_{0j}-\alpha_{1j}\right)p_{j}^{*}}+\frac{COV\left(Y,T|Z=j\right)}{\left(1-\alpha_{0j}-\alpha_{1j}\right)\left(1-p_{j}^{*}\right)}=\frac{COV\left(Y,T|Z=k\right)}{\left(1-\alpha_{0k}-\alpha_{1k}\right)p_{k}^{*}}+\frac{COV\left(Y,T|Z=k\right)}{\left(1-\alpha_{0k}-\alpha_{1k}\right)\left(1-p_{k}^{*}\right)}\\
\frac{1}{\left(1-\alpha_{0j}-\alpha_{1j}\right)}\cdot\frac{COV\left(Y,T|Z=j\right)}{p_{j}^{*}\left(1-p_{j}^{*}\right)}=\frac{1}{\left(1-\alpha_{0k}-\alpha_{1k}\right)}\cdot\frac{COV\left(Y,T|Z=k\right)}{p_{k}^{*}\left(1-p_{k}^{*}\right)}
\end{gathered}$$

where we also make use of the definition of p_j for the following substitutions $p_j-\alpha_{0j}=\left(1-\alpha_{0j}-\alpha_{1j}\right)p_j^{*}$ and $1-p_j-\alpha_{1j}=\left(1-\alpha_{0j}-\alpha_{1j}\right)\left(1-p_j^{*}\right)$.

A.4 Proof of Theorem 1

Let the instrument Z take on three values, $j=0,1,2$. Re-writing (2) in terms of β and then equating this result for the cases $(j=1,k=0)$ and $(j=2,k=1)$ yields

$$\frac{E\left[Y|Z=1\right]-E\left[Y|Z=0\right]}{p_{1}^{*}-p_{0}^{*}}=\frac{E\left[Y|Z=2\right]-E\left[Y|Z=1\right]}{p_{2}^{*}-p_{1}^{*}}$$

Letting $\Delta_{jk}=E\left[Y|Z=j\right]-E\left[Y|Z=k\right]$, we can solve this expression for p_1^{*}

$$p_{1}^{*}=\frac{p_{2}^{*}\Delta_{10}+p_{0}^{*}\Delta_{21}}{\Delta_{21}+\Delta_{10}} \tag{23}$$

which we can use to derive an expression for $p_1^*\left(1-p_1^*\right)$

$$\begin{aligned}
p_1^*\left(1-p_1^*\right) &= \left(\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\left(\frac{\Delta_{21}+\Delta_{10}}{\Delta_{21}+\Delta_{10}}-\left(\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\right)\\
&= \left(\frac{p_2^*\Delta_{10}+p_0^*\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\left(\frac{\left(1-p_2^*\right)\Delta_{10}+\left(1-p_0^*\right)\Delta_{21}}{\Delta_{21}+\Delta_{10}}\right)\\
&= \left(\frac{1}{\Delta_{20}}\right)^2\left(p_2^*\left(1-p_2^*\right)\Delta_{10}^2+p_0^*\left(1-p_0^*\right)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)\right)
\end{aligned} \tag{24}$$

where the last step uses $\Delta_{20}=\Delta_{21}+\Delta_{10}$. Next, we can re-write equation (8) for the case $j=1,k=0$ to yield

$$p_1^*\left(1-p_1^*\right)=p_0^*\left(1-p_0^*\right)\cdot\frac{C_1}{C_0} \tag{25}$$

where $C_j=COV\left(Y,T|Z=j\right)$. Inserting (24) into (25) and simplifying yields

$$\begin{aligned}
&\left(\frac{1}{\Delta_{20}}\right)^2\left(p_2^*\left(1-p_2^*\right)\Delta_{10}^2+p_0^*\left(1-p_0^*\right)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)\right)=p_0^*\left(1-p_0^*\right)\cdot\frac{C_1}{C_0}\\
&p_2^*\left(1-p_2^*\right)\Delta_{10}^2+p_0^*\left(1-p_0^*\right)\left(\Delta_{21}^2-\Delta_{20}^2\frac{C_1}{C_0}\right)+\Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)=0
\end{aligned} \tag{26}$$

Similarly, we can re-write equation (8) for the case $\left(j=2,k=1\right)$

$$p_1^*\left(1-p_1^*\right)=p_2^*\left(1-p_2^*\right)\cdot\frac{C_1}{C_2} \tag{27}$$

which we can combine with (24) to yield

$$\begin{aligned}
&\left(\frac{1}{\Delta_{20}}\right)^2\left(p_2^*\left(1-p_2^*\right)\Delta_{10}^2+p_0^*\left(1-p_0^*\right)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)\right)=p_2^*\left(1-p_2^*\right)\cdot\frac{C_1}{C_2}\\
&p_2^*\left(1-p_2^*\right)\left(\Delta_{10}^2-\Delta_{20}^2\frac{C_1}{C_2}\right)+p_0^*\left(1-p_0^*\right)\Delta_{21}^2+\Delta_{10}\Delta_{21}\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)=0
\end{aligned} \tag{28}$$

To find the solutions to the remaining two unknowns, p_0^* and p_2^* given the two equations (26) and (28), we proceed in the following steps. First, we re-write these two equations as quadratics in p_0^*. We can then compute the resultant using the coefficients on p_0^* in the re-written equations. As these coefficients are only a function of p_2^*, the resultant also is only a function of p_2^*. Since the system of equations has a non-zero solutions if and only if the resultant equals zero, we set the resultant equal to zero and solve for roots of p_2^*. We then repeat the process by re-writing equations (26) and (28) as quadratics in p_2^* and solve that resultant for roots of p_0^*. Finally, we substitute all combinations of the roots of p_0^* and p_2^* into (26) and (28) to find the pair(s) of roots that satisfy these equations.

Re-writing both (26) and (28) as quadratics in p_0^* yields

$$\begin{aligned}\left(\Delta_{20}^2\frac{C_1}{C_0}-\Delta_{21}^2\right)p_0^{*2}+\left(\Delta_{21}^2-\Delta_{20}^2\frac{C_1}{C_0}+\Delta_{10}\Delta_{21}-2\Delta_{10}\Delta_{21}p_2^*\right)p_0^*&\\ +\left(\left(\Delta_{10}^2+\Delta_{10}\Delta_{21}\right)p_2^*-\Delta_{10}^2p_2^{*2}\right)=0&\qquad(29)\\ -\Delta_{21}^2p_0^{*2}+\left(\Delta_{21}^2+\Delta_{10}\Delta_{21}-2\Delta_{10}\Delta_{21}p_2^*\right)p_0^*&\\ +\left(\left(\Delta_{10}^2-\Delta_{20}^2\frac{C_1}{C_2}+\Delta_{10}\Delta_{21}\right)p_2^*+\left(\Delta_{20}^2\frac{C_1}{C_2}-\Delta_{10}^2\right)p_2^{*2}\right)=0&\qquad(30)\end{aligned}$$

Denoting the coefficients of the system of quadratic equations formed by (29) and (30) by

$$\begin{aligned}A_{00}&=\Delta_{20}^2\frac{C_1}{C_0}-\Delta_{21}^2\\ A_{01}&=\Delta_{21}^2-\Delta_{20}^2\frac{C_1}{C_0}+\Delta_{10}\Delta_{21}-2\Delta_{10}\Delta_{21}p_2^*\\ A_{02}&=\left(\Delta_{10}^2+\Delta_{10}\Delta_{21}\right)p_2^*-\Delta_{10}^2p_2^{*2}\\ B_{00}&=-\Delta_{21}^2\\ B_{01}&=\Delta_{21}^2+\Delta_{10}\Delta_{21}-2\Delta_{10}\Delta_{21}p_2^*\\ B_{02}&=\left(\Delta_{10}^2-\Delta_{20}^2\frac{C_1}{C_2}+\Delta_{10}\Delta_{21}\right)p_2^*+\left(\Delta_{20}^2\frac{C_1}{C_2}-\Delta_{10}^2\right)p_2^{*2}\end{aligned}$$

we can form the Sylvester matrix, $S_{p_0^*}$, of this system of equations in p_0^*

$$\begin{bmatrix}A_{00}&A_{01}&A_{02}&0\\0&A_{00}&A_{01}&A_{02}\\B_{00}&B_{01}&B_{02}&0\\0&B_{00}&B_{01}&B_{02}\end{bmatrix}$$

Since the resultant can be formed as the determinant of the Sylvester matrix, taking the determinant of $S_{p_0^*}$ and setting it equal to zero yields

$$\begin{aligned}
&\frac{C_1^2\Delta_{20}^4\left(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2)\right)}{C_0^2C_2^2}\cdot {p_2^*}^4\\
&-\frac{2C_1^2\Delta_{20}^4\left(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2)\right)}{C_0^2C_2^2}\cdot {p_2^*}^3\\
&+\frac{1}{C_0^2C_2^2}\Big[C_1^2\Delta_{20}^4\left(C_0^2\Delta_{21}^4+C_1^2\Delta_{20}^4-C_0C_2\Delta_{10}\Delta_{21}^2(3\Delta_{20})+C_2^2\Delta_{10}^2(\Delta_{10}^2-\Delta_{10}\Delta_{21}-\Delta_{21}^2)\right.\\
&\left.-C_1\Delta_{20}^2(C_2\Delta_{10}(2\Delta_{10}-\Delta_{21})+2C_0\Delta_{21}^2)\right)\Big]\cdot {p_2^*}^2\\
&+\frac{C_1^2\Delta_{10}\Delta_{21}\Delta_{20}^5(C_2\Delta_{10}+C_0\Delta_{21}-C_1\Delta_{20})}{C_0^2C_2}\cdot p_2^*=0 \qquad (31)
\end{aligned}$$

Following an analogous process, we re-write (26) and (28) as quadratics in p_2^*, form the Sylvester matrix $S_{p_2^*}$, take its determinant and set it equal to zero to yield

$$\begin{aligned}
&\frac{C_1^2\Delta_{20}^4(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2))}{C_0^2C_2^2}\cdot {p_0^*}^4\\
&-\frac{2C_1^2\Delta_{20}^4\left(C_1^2\Delta_{20}^4+(C_2\Delta_{10}^2-C_0\Delta_{21}^2)^2-2C_1\Delta_{20}^2(C_2\Delta_{10}^2+C_0\Delta_{21}^2)\right)}{C_0^2C_2^2}\cdot {p_0^*}^3\\
&+\frac{1}{C_0^2C_2^2}\Big[C_1^2\Delta_{20}^4\left(C_2^2\Delta_{10}^4+C_1^2\Delta_{20}^4-C_0C_2\Delta_{10}^2\Delta_{21}(\Delta_{10}+3\Delta_{21})+C_0^2\Delta_{21}^2(-\Delta_{10}^2-\Delta_{10}\Delta_{21}+\Delta_{21}^2)\right.\\
&\left.-C_1\Delta_{20}^2(2C_2\Delta_{10}^2+C_0\Delta_{21}(-\Delta_{10}+2\Delta_{21}))\right)\Big]\cdot {p_0^*}^2\\
&+\frac{C_1^2\Delta_{10}\Delta_{21}\Delta_{20}^5(C_2\Delta_{10}+C_0\Delta_{21}-C_1\Delta_{20})}{C_0C_2^2}\cdot p_0^*=0 \qquad (32)
\end{aligned}$$

We find that (31) yields four roots of p_2^* and (32) yields four roots of p_0^*. We then use all combinations of these roots to determine which (p_0^*, p_2^*) root pairs satisfy (26) and (28).

Two of the four roots for both p_0^* and p_2^* are 0 and 1. When either p_0^* or p_2^* equals 0, the solution to the system of equations is $p_0^* = 0$ and $p_2^* = 0$. Similarly, when either p_0^* or p_2^* equals 1, the solution to the system of equations is $p_0^* = 1$ and $p_2^* = 1$. However, these two pairs of solutions violate Assumption 1.iii that $0 < p_j^* < 1,\ \forall j$ and $p_j^* \neq p_k^*,\ j \neq k$. Thus, neither 0 nor 1 is a possible solution for either p_0^* or p_2^*.

The two pairs of roots which satisfy (26) and (28) for which $p_j^* \neq p_k^*,\ j \neq k$, along with corresponding solution for p_1^* computed by inserting each (p_0^*, p_2^*) pair into (23), are

$$\begin{aligned} p_0^* &= \frac{D_3 + \sqrt{D_3}\Big((C_1 - C_0)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{10}^2\Big)}{2D_3} \\ p_1^* &= \frac{D_3 + \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_1 - C_0)\,\Delta_{21}^2\Big)}{2D_3} \\ p_2^* &= \frac{D_3 + \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{21}^2\Big)}{2D_3} \end{aligned} \tag{33}$$

and

$$\begin{aligned} p_0^* &= \frac{D_3 - \sqrt{D_3}\Big((C_1 - C_0)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{10}^2\Big)}{2D_3} \\ p_1^* &= \frac{D_3 - \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_1 - C_0)\,\Delta_{21}^2\Big)}{2D_3} \\ p_2^* &= \frac{D_3 - \sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{21}^2\Big)}{2D_3} \end{aligned} \tag{34}$$

where $D_0 = C_0\Delta_{21}^2$, $D_1 = C_1\Delta_{20}^2$, $D_2 = C_2\Delta_{10}^2$, and $D_3 = D_0^2 + D_1^2 + D_2^2 - 2D_0D_1 - 2D_0D_2 - 2D_1D_2$.

Both sets of solutions yield real numbers if $D_3 > 0$. We prove that is the case below in section A.4.1.

We can next solve for β by selecting any pair of p_j^* and substituting the appropriate values into equation (2). Without loss of generality, using the first solution found in equations (33) along with the fact that $\Delta_{20}^2 = (\Delta_{21} + \Delta_{10})^2 = \Delta_{21}^2 + \Delta_{10}^2 + 2\Delta_{21}\Delta_{10}$, we can first compute $p_1^* - p_0^*$

$$\begin{aligned} p_1^* - p_0^* &= \frac{\sqrt{D_3}\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_1 - C_0)\,\Delta_{21}^2\Big)}{2D_3} - \frac{\sqrt{D_3}\Big((C_1 - C_0)\,\Delta_{20}^2 + (C_0 - C_2)\,\Delta_{10}^2\Big)}{2D_3} \\ &= \frac{\sqrt{D_3}\Big((2C_2 - C_1 - C_0)\,\Delta_{10}^2 + (C_1 - C_0)\left(\Delta_{21}^2 - \Delta_{20}^2\right)\Big)}{2D_3} \\ &= \frac{\sqrt{D_3}\Big(2\,(C_2 - C_1)\,\Delta_{10}^2 + 2\,(C_0 - C_1)\,(\Delta_{21}\Delta_{10})\Big)}{2D_3} \\ &= \frac{\Big((C_2 - C_1)\,\Delta_{10}^2 + (C_0 - C_1)\,(\Delta_{21}\Delta_{10})\Big)}{\sqrt{D_3}} \end{aligned} \tag{35}$$

Thus, for the first solution found in equations (33), re-writing equation (2) to solve for β and then inserting the above result for $p_1^* - p_0^*$ yieids

$$\begin{aligned}\beta &= \frac{\Delta_{10}}{p_1^* - p_0^*} \\ &= \frac{\sqrt{D_3}\Delta_{10}}{\Big(\left(C_2 - C_1\right) \Delta_{10}^2 + \left(C_0 - C_1\right) \left(\Delta_{21}\Delta_{10}\right) \Big)} \\ &= \frac{\sqrt{D_3}}{\Big(\left(C_2 - C_1\right) \Delta_{10} + \left(C_0 - C_1\right) \Delta_{21} \Big)} \qquad (36)\end{aligned}$$

Similarly, we can show that β corresponding to the second solution found in equations (34), is equal and oppose in sign from the result in equation (36), i.e.,

$$\beta = -\frac{\sqrt{D_3}}{\Big(\left(C_2 - C_1\right) \Delta_{10} + \left(C_0 - C_1\right) \Delta_{21} \Big)} \qquad (37)$$

To solve for $\overline{\alpha}$, notice that under the Assumption 2, an OLS regression of Y on T^* will yield a consistent estimate of β. Moreover, a regression of Y on T^* conditional on any value taken on by the instrument will also consistently estimate β. Therefore,

$$\beta = \frac{COV\left(Y, T^* | Z = j\right)}{VAR\left(T^* | Z = j\right)} = \frac{COV\left(Y, T^* | Z = j\right)}{p_j^* \left(1 - p_j^*\right)} = \frac{COV\left(Y, T | Z = j\right)}{\left(1 - \alpha_{0j} - \alpha_{1j}\right) p_j^* \left(1 - p_j^*\right)} \qquad (38)$$

where the last equality results from the fact that $COV\left(Y, T | Z = j\right) = \left(1 - \alpha_{0j} - \alpha_{1j}\right) COV\left(Y, T^* | Z = j\right)$ which we prove below in section A.4.2.

Re-writing (38) to solve for $\overline{\alpha} = \alpha_{0j} + \alpha_{1j}$ yields

$$1 - \overline{\alpha} = \frac{COV\left(Y, T | Z = j\right)}{\beta p_j^* \left(1 - p_j^*\right)} \qquad (39)$$

Inspection of equations (33) and (34) reveals that the results for each p_j^* sum to one across the two solutions, i.e., p_j^* in (33) equals $1 - p_j^*$ in (34). Thus, $p_j^* \left(1 - p_j^*\right)$ is the same across both solutions as is the observed quantity $COV\left(Y, T | Z = j\right)$. Since the corresponding results for β are equal but opposite in sign across the two solutions, the right-hand side of (39) is positive for one solution and negative for another solution. Thus $\overline{\alpha} < 1$ for one solution and $\overline{\alpha} > 1$ for the other solution. Since Assumption 1.v requires $\overline{\alpha} < 1$, we can determine which of the two solutions matches this assumption and therefore determine the sign of β.

A.4.1 Proof that $D_3 > 0$

The solutions in equations (33) and (34) require $D_3 > 0$ in order to yield real numbers. Rewriting the expression for $p_1^* (1 - p_1^*)$, found in (24), making substitutions using equation (8), and simplifying yields

$$\begin{aligned}
p_1^* (1 - p_1^*) \Delta_{20}^2 &= p_2^* (1 - p_2^*) \Delta_{10}^2 + p_0^* (1 - p_0^*) \Delta_{21}^2 + \Delta_{10}\Delta_{21} (p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*)) \\
\left[p_0^* (1 - p_0^*) \cdot \frac{C_1}{C_0}\right] \Delta_{20}^2 &= \left[p_0^* (1 - p_0^*) \cdot \frac{C_2}{C_0}\right] \Delta_{10}^2 + p_0^* (1 - p_0^*) \Delta_{21}^2 + \Delta_{10}\Delta_{21} (p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*)) \\
C_1\Delta_{20}^2 &= C_2\Delta_{10}^2 + C_0\Delta_{21}^2 + C_0\Delta_{10}\Delta_{21} \frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))}{p_0^* (1 - p_0^*)} \\
D_1 - D_2 - D_0 &= C_0\Delta_{10}\Delta_{21} \frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))}{p_0^* (1 - p_0^*)} \qquad (40)
\end{aligned}$$

Squaring both sides of equation (40) and simplifying yields

$$\begin{aligned}
(D_1 - D_2 - D_0)^2 &= C_0^2\Delta_{10}^2\Delta_{21}^2 \frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))^2}{(p_0^* (1 - p_0^*))^2} \\
D_0^2 + D_1^2 + D_2^2 - 2D_0D_1 - 2D_1D_2 + 2D_0D_2 &= \left[\frac{C_0}{p_0^* (1 - p_0^*)}\Delta_{10}^2\right] C_0\Delta_{21}^2 \frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))^2}{p_0^* (1 - p_0^*)} \\
D_3 + 4D_0D_2 &= \left[\frac{C_2}{p_2^* (1 - p_2^*)}\Delta_{10}^2\right] C_0\Delta_{21}^2 \frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))^2}{p_0^* (1 - p_0^*)} \\
D_3 + 4D_0D_2 &= D_2D_0 \frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))^2}{p_0^* (1 - p_0^*) p_2^* (1 - p_2^*)} \\
D_3 &= D_0D_2 \left[\frac{(p_2^* (1 - p_0^*) + p_0^* (1 - p_2^*))^2}{p_0^* (1 - p_0^*) p_2^* (1 - p_2^*)} - 4\right] \qquad (41)
\end{aligned}$$

Notice that C_0 and C_2 must always have the same sign in order to satisfy equation (8). Thus, $C_0C_2 > 0$ and, in turn, $D_0D_2 = C_0\Delta_{21}^2C_2\Delta_{10}^2 > 0$.[39] Therefore, $D_3 > 0$ as long as the term in

[39] Rule out $C_j = 0 \ \forall \ j$.

brackets on the right hand side of (41) is positive. To confirm this condition holds notice that

$$\begin{aligned}
\frac{\left(p_2^*\left(1-p_0^*\right)+p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)}-4&>0\\
\frac{\left(p_2^*\left(1-p_0^*\right)\right)^2+\left(p_0^*\left(1-p_2^*\right)\right)^2+2p_2^*\left(1-p_0^*\right)p_0^*\left(1-p_2^*\right)}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)}&>4\\
\frac{\left(p_2^*\left(1-p_0^*\right)\right)^2+\left(p_0^*\left(1-p_2^*\right)\right)^2}{p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)}+2&>4\\
\left(p_2^*\left(1-p_0^*\right)\right)^2+\left(p_0^*\left(1-p_2^*\right)\right)^2&>2p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)\\
\left(p_2^*\left(1-p_0^*\right)\right)^2-2p_0^*\left(1-p_0^*\right)p_2^*\left(1-p_2^*\right)+\left(p_0^*\left(1-p_2^*\right)\right)^2&>0\\
\left[\left(p_2^*\left(1-p_0^*\right)\right)-\left(p_0^*\left(1-p_2^*\right)\right)\right]^2&>0\\
\left[p_2^*-p_0^*\right]^2&>0\\
\frac{\Delta_{20}^2}{\beta^2}&>0 \qquad (42)
\end{aligned}$$

As long as $\beta \neq 0$, this condition is satisfied and, thus, $D_3 > 0$.

A.4.2 Proof that $COV\left(W,T|Z=j\right)=\left(1-\alpha_{0j}-\alpha_{1j}\right)COV\left(W,T^*|Z=j\right)$

For a given variable, W, the covariance between W and T^* among those observations where $Z = j$ simplifies to

$$\begin{aligned}
COV\left(W,T^*|Z=j\right)&=E\left[WT^*|Z=j\right]-E\left[W|Z=j\right]E\left[T^*|Z=j\right]\\
&=p_j^*E\left[W|T^*=1,Z=j\right]-E\left[W|Z=j\right]p_j^*\\
&=\big(E\left[W|T^*=1,Z=j\right]-E\left[W|Z=j\right]\big)p_j^* \qquad (43)
\end{aligned}$$

where we make use of the fact that

$$E\left[WT^*|Z=j\right]=p_j^*E\left[W\cdot 1|T^*=1,Z=j\right]+\left(1-p_j^*\right)E\left[W\cdot 0|T^*=0,Z=j\right]=p_j^*E\left[W|T^*=1,Z=j\right]$$

Similarly, for a given variable, W, the covariance between W and T among those observations where $Z = j$ simplifies to

$$COV\left(W, T|Z = j\right) = \big(E\left[W|T = 1, Z = j\right] - E\left[W|Z = j\right]\big)p_j \tag{44}$$

Notice that replacing Y with W in equation (19) and extending the assumption of non-differential measurement error to W yields an expression for $E\left[W|T = 1, Z = j\right]$. Inserting this term into equation (44) yields

$$\begin{aligned}
COV\left(W, T|Z = j\right) = &\Bigg(E\left[W|T^* = 0, Z = j\right] \cdot \frac{\alpha_{0j}\left(1 - p_j^*\right)}{p_j} + E\left[W|T^* = 1, Z = j\right] \cdot \frac{\left(1 - \alpha_{1j}\right)p_j^*}{p_j} \\
&- E\left[W|Z = j\right]\Bigg)p_j \\
&= E\left[W|T^* = 0, Z = j\right] \cdot \alpha_{0j}\left(1 - p_j^*\right) + E\left[W|T^* = 1, Z = j\right] \cdot \left(1 - \alpha_{1j}\right)p_j^* \\
&- E\left[W|Z = j\right]p_j
\end{aligned} \tag{45}$$

Finally, adding zero in the form of $\alpha_{0j}E\left[W|Z = j\right] - \alpha_{0j}E\left[W|Z = j\right]$ yields

$$\begin{aligned}
COV\left(W, T|Z = j\right) = &E\left[W|T^* = 0, Z = j\right] \cdot \alpha_{0j}\left(1 - p_j^*\right) + E\left[W|T^* = 1, Z = j\right] \cdot \left(1 - \alpha_{1j}\right)p_j^* \\
&- \alpha_{0j}E\left[W|Z = j\right] \\
&- \big(E\left[W|Z = j\right]p_j - \alpha_{0j}E\left[W|Z = j\right]\big) \\
&= E\left[W|T^* = 0, Z = j\right] \cdot \alpha_{0j}\left(1 - p_j^*\right) + E\left[W|T^* = 1, Z = j\right] \cdot \left(1 - \alpha_{1j}\right)p_j^* \\
&- \alpha_{0j}\Big(E\left[W|T^* = 0, Z = j\right]\left(1 - p_j^*\right) + E\left[W|T^* = 1, Z = j\right]p_j^*\Big) \\
&- E\left[W|Z = j\right]\left(p_j - \alpha_{0j}\right) \\
&= E\left[W|T^* = 1, Z = j\right] \cdot \left(1 - \alpha_{0j} - \alpha_{1j}\right)p_j^* \\
&- E\left[W|Z = j\right] \cdot \left(1 - \alpha_{0j} - \alpha_{1j}\right)p_j^* \\
&= \big(E\left[W|T^* = 1, Z = j\right] - E\left[W|Z = j\right]\big) \cdot \left(1 - \alpha_{0j} - \alpha_{1j}\right)p_j^* \\
&= \left(1 - \alpha_{0j} - \alpha_{1j}\right)COV\left(W, T^*|Z = j\right)
\end{aligned} \tag{46}$$

A.5 Including covariates

Including covariates requires the following modifications to Assumptions 1, 2, and 4 to account for the additional regressor(s).

Assumption 1′.ii: $E[Y|Z=j,\boldsymbol{X}] = \gamma + \beta p_j^* + E[\boldsymbol{\psi X}|Z=j,\boldsymbol{X}],\ \forall j$

Assumption 1′.iv: $E\left[Y|T,T^*,Z,\boldsymbol{X}\right] = E\left[Y|T^*,Z,\boldsymbol{X}\right]$

Assumption 2′: $E\left[Y|T^*=t,Z=j,\boldsymbol{X}\right] = E\left[Y|T^*=t,\boldsymbol{X}\right],\ \forall j$

Assumption 4′: $E\left[Y|T^*=1,Z=j,\boldsymbol{X}\right] - E\left[Y|T^*=0,Z=j,\boldsymbol{X}\right] = E\left[Y|T^*=1,Z=k,\boldsymbol{X}\right] - E\left[Y|T^*=0,Z=k,\boldsymbol{X}\right]\ \ \forall\ j,k$

We also assume that each element in $\boldsymbol{X}$ is uncorrelated with the error ϵ. The analogous covariance equations to (8) are

$$\frac{COV\left(Y,T|Z=j\right)}{p_j^*\left(1-p_j^*\right)} - \frac{COV\left(\boldsymbol{\psi X},T|Z=j\right)}{p_j^*\left(1-p_j^*\right)} = \frac{COV\left(Y,T|Z=k\right)}{p_k^*\left(1-p_k^*\right)} - \frac{COV\left(\boldsymbol{\psi X},T|Z=k\right)}{p_k^*\left(1-p_k^*\right)} \tag{47}$$

where we again eliminate the misclassification rates from the system of moment conditions by assuming that the sum of the misclassification rates is constant across instrument values. We can obtain (47) using $COV\left(Y,T^*|Z=j\right)$ since

$$\begin{aligned}
COV\left(Y,T^*|Z=j\right) &= \beta COV\left(T^*,T^*|Z=j\right) + COV\left(\boldsymbol{\psi X},T^*|Z=j\right) + COV\left(\epsilon,T^*|Z=j\right)\\
COV\left(Y,T^*|Z=j\right) &= \beta p_j^*\left(1-p_j^*\right) + COV\left(\boldsymbol{\psi X},T^*|Z=j\right)\\
\frac{COV\left(Y,T^*|Z=j\right)}{p_j^*\left(1-p_j^*\right)} &= \beta + \frac{COV\left(\boldsymbol{\psi X},T^*|Z=j\right)}{p_j^*\left(1-p_j^*\right)}\\
\frac{COV\left(Y,T|Z=j\right)}{\left(1-\alpha_{0j}-\alpha_{1j}\right)p_j^*\left(1-p_j^*\right)} &= \beta + \frac{COV\left(\boldsymbol{\psi X},T|Z=j\right)}{\left(1-\alpha_{0j}-\alpha_{1j}\right)p_j^*\left(1-p_j^*\right)}\\
\frac{COV\left(Y,T|Z=j\right)}{\left(1-\overline{\alpha}\right)p_j^*\left(1-p_j^*\right)} &= \beta + \frac{COV\left(\boldsymbol{\psi X},T|Z=j\right)}{\left(1-\overline{\alpha}\right)p_j^*\left(1-p_j^*\right)} \qquad (48)
\end{aligned}$$

The second line makes use of Assumption 2 (that T^* is an exogenous covariate), the fourth line makes use of $COV\left(W,T|Z=j\right) = \left(1-\alpha_{0j}-\alpha_{1j}\right)COV\left(W,T^*|Z=j\right)$ (see section A.4.2), and the final line makes use of Assumption 3. For any two instrument values, $Z = j$ and $Z = k$, we can then difference equation (48) and simplify to obtain the desired result.

The analog to equation (10), which can be obtained from re-arranging the terms of (48), is

$$1 - \overline{\alpha} = \frac{COV\left(Y, T|Z = j\right) - COV\left(\boldsymbol{\psi}\boldsymbol{X}, T|Z = j\right)}{\beta p_j^* \left(1 - p_j^*\right)} \tag{49}$$

As we note in the text, we can pre-estimate $\boldsymbol{\psi}$, the coefficient on $\boldsymbol{X}$, by regressing Y on T and $\boldsymbol{X}$ while using Z as an instrument for T. The proof that $\boldsymbol{\psi}$ can be consistently estimated in this way follows from a standard textbook analysis of using T as a proxy variable for T^*, but with one modification. Specifically, the key assumptions necessary for using T as a proxy variable are typically motivated using the linear projection of T^* on T

$$T^* = \lambda_0 + \lambda_1 T + r$$

to re-write the structural equation (16) as follows

$$Y = (\gamma + \beta\lambda_0) + \beta\lambda_1 T + \boldsymbol{\psi}\boldsymbol{X} + (\beta r + \epsilon) \tag{50}$$

The standard proxy variable method would then use OLS to estimate (50) based on the assumption that the composite error term $(\beta r + \epsilon)$ is uncorrelated with T and $\boldsymbol{X}$. The first assumption necessary for this method to be valid is that T be "redundant" in the structural equation. The definition of redundancy in Wooldridge (2010; see (4.25), p. 68) is equivalent to our assumption of non-differential measurement error (see Assumption $1'$.iv), which satisfies the first proxy variable assumption.

The second assumption that must be satisfied to use OLS is that $COV(r, \boldsymbol{X}) = 0$.[40] Rather than make this additional assumption, we instead make use of that fact we have already assumed an instrument exists for T, and instead estimate (50) using two-stage least squares. Using $\widehat{T}$ rather than T in (50) imposes that $COV(r, \boldsymbol{X}) = 0$ because $\boldsymbol{X}$ is included in the construction of $\widehat{T}$. It is

[40] The additional assumption requires that T^* be uncorrelated with $\boldsymbol{X}$ after we partial out T, which is given in (4.26) of Wooldridge (2010).

worth noting that the coefficient on $\widehat{T}$, the predicted proxy variable, is not a consistent estimator for β, but instead is a consistent estimator for $\beta\lambda_1$.

A.6 Identification of $\overline{\alpha}$ Under Assumption 4 Using Homoskedasticity

One possibility for identifying $\overline{\alpha}$ when using Assumption 4 is to impose restrictions on the higher order moments. In particular, we can require that the error term, ϵ, in equation (1) is homoskedastic, i.e., $E\left[\epsilon^2|Z=j\right] = E\left[\epsilon^2\right]$.

Following equation (1), the expected value of the square of Y when $Z=j$ is

$$\begin{aligned}
E\left[Y^2|Z=j\right] &= E\left[(\gamma+\beta T^*+\epsilon)^2|Z=j\right] \\
&= E\left[\gamma^2+\beta^2(T^*)^2+\epsilon^2+2\beta T^*\gamma+2\gamma\epsilon+2\beta T^*\epsilon|Z=j\right] \\
&= \beta^2 E\left[(T^*)^2|Z=j\right]+2\beta E\left[(\gamma+\epsilon)T^*|Z=j\right]+E\left[\gamma^2+2\gamma\epsilon+\epsilon^2|Z=j\right] \qquad (51)
\end{aligned}$$

The first term in (51) simplifies using $\beta^2 E\left[(T^*)^2|Z=j\right] = \beta^2 p_j^*$. The third term in (51) can be simplified using $E\left[\gamma^2+2\gamma\epsilon+\epsilon^2|Z=j\right] = \gamma^2+2\gamma E\left[\epsilon|Z=j\right]+E\left[\epsilon^2|Z=j\right] = \gamma^2+E\left[\epsilon^2\right]$ where the second equality follows both from the fact that $E\left[\epsilon|Z=j\right]$ and by the homoskedasticity assumption.

To find an expression for the second term in (51), we begin with right-hand side of (6) multiplied by $-\left(1-p_j^*\right)$. Simplifying this expression yields

$$\begin{aligned}
\left(1-p_j^*\right)\left[\frac{COV\left(Y,T|Z=j\right)}{1-p_j-\alpha_{1j}}-E\left[Y|Z=j\right]\right] &= \frac{COV\left(Y,T|Z=j\right)}{1-\alpha_{0j}-\alpha_{1j}}-\left(1-p_j^*\right)E\left[Y|Z=j\right] \\
&= COV\left(Y,T^*|Z=j\right)-\left(1-p_j^*\right)E\left[Y|Z=j\right] \\
&= \left(E\left[YT^*|Z=j\right]-E\left[Y|Z=j\right]p_j^*\right)-\left(1-p_j^*\right)E\left[Y|Z=j\right] \\
&= E\left[(\gamma+\beta T^*+\epsilon)T^*|Z=j\right]-E\left[Y|Z=j\right] \\
&= E\left[(\gamma+\epsilon)T^*|Z=j\right]+E\left[\beta(T^*)^2|Z=j\right]-E\left[Y|Z=j\right] \\
&= E\left[(\gamma+\epsilon)T^*|Z=j\right]+\beta p_j^*-\left(\gamma+\beta p_j^*\right) \\
&= E\left[(\gamma+\epsilon)T^*|Z=j\right]-\gamma \qquad (52)
\end{aligned}$$

where the first equality uses the fact that $1 - p_j - \alpha_{1j} = (1 - \alpha_{0j} - \alpha_{1j})\left(1 - p_j^*\right)$ and the second equality uses the result $COV\,(Y, T|Z = j) = (1 - \alpha_{0j} - \alpha_{1j})\,COV\,(Y, T^*|Z = j)$.

Substituting these results back into (51) yields

$$E\left[Y^2|Z = j\right] = \beta^2 p_j^* + 2\beta\left[\frac{COV\,(Y, T|Z = j)}{1 - \alpha_{0j} - \alpha_{1j}} - \left(1 - p_j^*\right)E\left[Y|Z = j\right] + \gamma\right] + \gamma^2 + E\left[\epsilon^2\right] \quad (53)$$

Differencing (53) for any two values of the instrument, j and k, yields

$$\begin{aligned} E\left[Y^2|Z = j\right] - E\left[Y^2|Z = k\right] &= \beta^2\left(p_j^* - p_k^*\right) + 2\beta\left[\frac{COV\,(Y, T|Z = j)}{1 - \alpha_{0j} - \alpha_{1j}} - \frac{COV\,(Y, T|Z = k)}{1 - \alpha_{0k} - \alpha_{1k}}\right] \\ &\quad - 2\beta\left[\left(1 - p_j^*\right)E\left[Y|Z = j\right] - \left(1 - p_k^*\right)E\left[Y|Z = k\right]\right] \\ &= \beta\left(E\left[Y|Z = j\right] - E\left[Y|Z = k\right]\right) \\ &\quad + 2\beta\left[\frac{COV\,(Y, T|Z = j)}{1 - \alpha_{0j} - \alpha_{1j}} - \frac{COV\,(Y, T|Z = k)}{1 - \alpha_{0k} - \alpha_{1k}}\right] \\ &\quad - 2\beta\left[\left(1 - p_j^*\right)E\left[Y|Z = j\right] - \left(1 - p_k^*\right)E\left[Y|Z = k\right]\right] \end{aligned} \quad (54)$$

where the second equality arises by using equation (2) to adjust the first term of (54). Again invoking Assumption 3 (i.e., $\overline{\alpha} = \alpha_{0j} + \alpha_{1j}$) and solving for $1 - \overline{\alpha}$ yields

$$1 - \overline{\alpha} = \frac{2\beta\left(COV\,(Y, T|Z = j) - COV\,(Y, T|Z = k)\right)}{\left(E\left[Y^2|Z = j\right] - E\left[Y^2|Z = k\right]\right) + \beta\left\{\Delta_{jk} + 2\left(E\left[Y|Z = j\right]p_j^* - E\left[Y|Z = k\right]p_k^*\right)\right\}} \quad (55)$$

where $\Delta_{jk} = E\left[Y|Z = j\right] - E\left[Y|Z = k\right]$.

To see that (55) implies that $\overline{\alpha} < 1$ for one solution and $\overline{\alpha} > 1$ for the other solution, let $\{\beta, p_j^*, p_k^*\}$ be the estimators for the first solution while $\{\tilde{\beta}, \tilde{p}_j^*, \tilde{p}_k^*\}$ are the corresponding estimators for the second solution. Recall from (9) that $\tilde{\beta} = -\beta$ and, since the estimators for a p_j^* sum to one across the two solutions, $\tilde{p}_j^* = 1 - p_j^*$ for each j.

Inserting these results into the second term in the denominator of (55) yields

$$
\begin{aligned}
\tilde{\beta}\left\{\Delta_{jk} + 2\left(E\left[Y|Z=j\right]\tilde{p}_j^* - E\left[Y|Z=k\right]\tilde{p}_k^*\right)\right\} &= -\beta\left\{E\left[Y|Z=j\right] - E\left[Y|Z=k\right]\right. \\
&\quad \left. + 2\left(E\left[Y|Z=j\right]\left(1-p_j^*\right) - E\left[Y|Z=k\right]\left(1-p_k^*\right)\right)\right\} \\
&= -\beta\left\{-\Delta_{jk} - 2\left(E\left[Y|Z=j\right]p_j^* - E\left[Y|Z=k\right]p_k^*\right)\right\} \\
&= \beta\left\{\Delta_{jk} + 2\left(E\left[Y|Z=j\right]p_j^* - E\left[Y|Z=k\right]p_k^*\right)\right\}
\end{aligned}
\tag{56}
$$

Thus, the second term in the denominator of (55) is the same for both solutions and, since $E\left[Y^2|Z=j\right] - E\left[Y^2|Z=k\right]$ does not vary across solutions, we see the entire denominator does not vary across the two solutions. The sign of the numerator, however, will vary across the two solutions since β is of equal magnitude but opposite in sign across the two solutions. Thus, $\overline{\alpha} < 1$ for one solution and $\overline{\alpha} > 1$ for the other solution.

A.7 Proof of Proposition 1

Since $COV\left(Y,T|Z=j\right) = \left(1-\alpha_{0j}-\alpha_{1j}\right)COV\left(Y,T^*|Z=j\right)$ (see section A.4.2) and $\alpha_{0j}+\alpha_{1j} < 1$ (by Assumption 1.v), it immediately follows that the sign of $\hat{\beta}_j^{OLS} = COV\left(Y,T|Z=j\right)/p_j\left(1-p_j\right)$ will be the same as $\beta = COV\left(Y,T^*|Z=j\right)/p_j^*\left(1-p_j^*\right)$.

Suppose $\beta > 0$ in which case $E\left[Y|T^*=1\right] > E\left[Y|T^*=0\right]$. Notice that for each instrument value j,

$$
\begin{aligned}
E\left[Y|T=1,Z=j\right] &= E\left[Y|T^*=0,T=1,Z=j\right]\cdot P[T^*=0|T=1,Z=j] \\
&\quad + E\left[Y|T^*=1,T=1,Z=j\right]\cdot P[T^*=1|T=1,Z=j] \\
&= E\left[Y|T^*=0\right]\cdot P[T^*=0|T=1,Z=j] \\
&\quad + E\left[Y|T^*=1\right]\cdot P[T^*=1|T=1,Z=j] \\
&\leq E\left[Y|T^*=1\right]\cdot P[T^*=0|T=1,Z=j] \\
&\quad + E\left[Y|T^*=1\right]\cdot P[T^*=1|T=1,Z=j] \\
&= E\left[Y|T^*=1\right]
\end{aligned}
\tag{57}
$$

where the second equality follows both from the non-differential measurement error assumption (Assumption 1.iv) and from Assumption 2 while the inequality follows from $\beta > 0$. Note that the inequality is strict if $P[T^* = 0|T = 1, Z = j] > 0$.

In addition

$$\begin{aligned} E\left[Y|T=1\right] &= \sum_j E\left[Y|T=1, Z=j\right] \cdot P[Z=j] \\ &\leq \sum_j \max_Z \{E\left[Y|T=1, Z=j\right]\} \cdot P[Z=j] \\ &= \max_Z \{E\left[Y|T=1, Z=j\right]\} \sum_j P[Z=j] \\ &= \max_Z \{E\left[Y|T=1, Z=j\right]\} \end{aligned} \tag{58}$$

where $\max_Z\{E\left[Y|T=1, Z=j\right]\}$ is the maximum value of $E\left[Y|T=1, Z=j\right]$ across all the instrument values and $\sum_j P[Z=j] = 1$.

Combining the results in (57) and (58), we see that

$$E\left[Y|T=1\right] \leq \max_Z \{E\left[Y|T=1, Z=j\right]\} \leq E\left[Y|T^*=1\right] \tag{59}$$

Analogously, it is straightforward to show that

$$E\left[Y|T=0\right] \geq \min_Z \{E\left[Y|T=0, Z=j\right]\} \geq E\left[Y|T^*=0\right] \tag{60}$$

Defining $\hat{\beta}^{LB,+} = \max_Z\{E\left[Y|T=1, Z=j\right]\} - \min_Z\{E\left[Y|T=0, Z=j\right]\}$, combining these last two results yields

$$\begin{gathered} E\left[Y|T=1\right] - E\left[Y|T=0\right] \leq \hat{\beta}^{LB,+} \leq E\left[Y|T^*=1\right] - E\left[Y|T^*=0\right] \\ 0 < \hat{\beta}^{OLS} \leq \hat{\beta}^{LB,+} \leq \beta \end{gathered} \tag{61}$$

where the last line uses the result that the sign of β^{OLS} is the same as the sign of β.

If $\beta < 0$, we define $\hat{\beta}^{LB,-} = \min_Z\{E\left[Y|T=1, Z=j\right]\} - \max_Z\{E\left[Y|T=0, Z=j\right]\}$ and using arguments analogous to those above, we can show that

$$0 > \hat{\beta}^{OLS} \geq \hat{\beta}^{LB,-} \geq \beta \tag{62}$$

A.8 Comparison to Other Estimators

In Appendix Table A4, we examine how other proposed estimators perform in the face of misclassification rates that vary. Columns (1) and (2) of Appendix Table A4 use exactly the same data generating process as columns (1) and (2) of Table 2, respectively. All of the estimators are well-centered in column (1) when the underlying misclassification rates are fixed. When we allow the underlying misclassifications rates to vary (but continue to assume that the sum is fixed) in column (2), the three estimators that rely on fixed misclassification and an instrumental variable (Frazis and Lowenstein 2003, Mahajan 2006, and Lewbel 2007) are not well-centered. The three estimators that rely on higher-order moments of the error term (Chen, Hu, and Lewbel 2008a; Chen, Hu, and Lewbel 2008b; DiTraglia and Garcia-Jimeno 2019) continue to be well-centered because the error term specification (normal errors) for the simulations in Table 2 are consistent with the higher order error term requirements of these estimators.

To examine the sensitivity of the estimators to these distributional assumptions, columns (3) through (5) of Appendix Table A4 are created to be identical to column (2), but differ in their specification for the error term. Specifically, consider the expanded DGP

$$Y = \gamma + \beta T^* + \epsilon_0(1 - T^*) + \epsilon_1 T^* \tag{63}$$

Letting $\epsilon_n \sim N(0, .25)$, we specify the error term as follows:

- Column (1): $\epsilon_0 = \epsilon_1 = \epsilon_n$
- Column (2): $\epsilon_0 = \epsilon_1 = \epsilon_n$
- Column (3): $\epsilon_0 = \epsilon_n$; $\epsilon_1 = 2.5\epsilon_n$
- Column (4): $\epsilon_0 = \epsilon_1 = |2.5\epsilon_n| - \overline{|2.5\epsilon_n|}$

- Column (5): $\epsilon_0 = |\epsilon_n| - \overline{|\epsilon_n|}$; $\epsilon_1 = |2.5\epsilon_n| - \overline{|2.5\epsilon_n|}$

Chen, Hu, and Lewbel (2008a) and DiTraglia and Garcia-Jimeno (2019) assume homoskedastic errors, so these estimators are not well-centered in column (3) that specifies a heteroskedastic error term.[41] Chen, Hu, and Lewbel (2008b) assumes that the error distribution has a third moment that is equal to zero, so this estimator starts is less well-centered in column (4) that specifies the error term to be a half-normal distribution. All three estimators are poorly centered in the final column that specifies a heteroskedastic and non-symmetric error term. For all columns in Appendix Table A4, our estimator is well-centered.

[41] Chen, Hu, and Lewbel (2008a) require that the errors are homoskedastic with respect to T^* while DiTraglia and Garcia-Jimeno (2019) require homoskedasticity with respect to Z. In addition, both estimators make a constant skewness assumption, with Chen, Hu, and Lewbel (2008a) requiring it with respect to T^* and with DiTraglia and Garcia-Jimeno (2019) requiring it with respect to Z.